

РЕЦЕНЗИИ / REVIEWS

M. Baerman, D. Brown, G. G. Corbett (eds.). Understanding and measuring morphological complexity. Oxford: Oxford University Press, 2015. 240 p. ISBN 978-0-198-72376-9.

Александр Бердичевский

Aleksandrs Berdicevskis

Упсальский университет, Упсала, 75105, Швеция;
Норвежский арктический университет, Тромсё,
9037, Норвегия;
aleksandrs.berdicevskis@lingfil.uu.se

Uppsala universitet, Uppsala, 75105, Sweden; UiT
Norges arktiske universitet, Tromsø, 9037, Norway;
aleksandrs.berdicevskis@lingfil.uu.se

Языковая сложность (в первую очередь морфологическая сложность) — чрезвычайно популярная тема в современной лингвистике, см. обзор ключевых зарубежных работ в [Бердичевский 2012], а также рецензию Н. Б. Вахтина [2014] на книгу П. Традгила. За пять лет, прошедших с момента публикации обзора, появилось много новых работ, среди них — рецензируемый сборник «Понимание и измерение морфологической сложности».

Двое из трех его редакторов (Мэтью Баерман и Гревил Корбет) — постоянные члены Суррейской морфологической группы, третий (Данстан Браун) с ней тесно связан. В обширный круг интересов группы уже давно входит и морфологическая сложность: этой теме, в частности, был посвящен шестилетний типологический проект, поддержанный Европейским исследовательским советом, а также ряд проектов чуть меньшего масштаба. Сборник написан по итогам конференции, состоявшейся в 2012 г. в Лондоне в рамках вышеупомянутого проекта.

Сборник состоит из трех частей: «Что такое морфологическая сложность?», «Понимание сложности» и «Измерение сложности». Две главы первой части содержат в основном общие рассуждения о природе морфологической сложности, четыре главы второй части — анализ морфологии конкретных языков (во всех случаях чрезвычайно сложной), четыре главы третьей части посвящены количественным подходам к сложности, хотя и не всегда напрямую ее измерению. Наибольшее внимание я уделяю третьей части, потому что она представляется мне наиболее интересной и потому что главы второй части содержат детальный анализ конкретных сложнейших парадигм, который плохо сжимается в пересказ.

Первая глава написана тремя редакторами-составителями и называется «Понимание и измерение морфологической сложности: введение». Авторы отмечают, что морфология (здесь и далее под «морфологией», если не оговорено иное, будет пониматься словоизменение, а не словообразование) резко отличается от других свойств языка и на макро-, и на микроуровне. Отличие на макроуровне состоит в том, что морфология — необязательный компонент. Языки без словоизменения встречаются, в то время как языков без синтаксиса, фонологии или семантики пока не обнаружено. Учитывая, что именно морфология может достигать исключительной сложности (по мнению авторов, вероятно, большей, чем иные компоненты), факт ее необязательности заслуживает внимания. Отличие на микроуровне состоит в том, что морфология наиболее идиосинкратична. Можно найти два неродственных языка, у которых похожи синтаксис и фонология, но сделать то же для словоизменения чрезвычайно трудно.

Эти два факта необходимо иметь в виду для адекватного описания словоизменяющих систем, чему примерно и посвящены остальные главы. Авторы также отмечают, что книга

рассматривает морфологию как отдельную структуру, не сводимую к синтаксису и фонологии, и концентрируется в основном на парадигматической, а не синтагматической сложности, то есть на сложности отношений между элементами парадигмы, а не на сложности объединения элементов в линейные последовательности (конструкции, предложения и т. п.).

По мнению авторов, главы книги демонстрируют, что словоизменительные системы могут быть устроены чрезвычайно затейливо, но тем не менее они обладают некоторой элегантностью и при достаточно абстрактном взгляде на них можно увидеть сравнимые структуры. Что именно авторы имеют в виду, для меня, увы, остается не вполне ясным. Главы второй части содержат много подробной информации о конкретных языках, а также наблюдений, важных для более общих теоретических и типологических исследований сложности; главы третьей части описывают ряд перспективных количественных подходов, но я не могу сказать, что мне как читателю удалось за восемью эмпирическими главами разглядеть некоторую общую «элегантность» морфологии и особенно «сравнимые структуры» описываемых языков.

В заключение авторы замечают, что книга оставляет открытым вопрос, почему же эти сложность и элегантность некоторым языкам присущи в высочайшей степени, а в других полностью отсутствуют. Это очень важная оговорка, к которой мы вернемся в конце рецензии.

Вторая глава написана Стивеном Андерсоном (Stephen Anderson) и называется «Dimensions of morphological complexity». Буквально название переводится как «Измерения морфологической сложности», но я бы назвал главу по-русски «Возможные типы морфологической сложности и ее составляющие», потому что описывает автор именно это — те свойства морфологии, которые делают ее «сложной». Начиная с попытки определить сложность, Андерсон приводит поучительный пример: он вел курс о языках мира в Пекинском университете. Студенты были знакомы с понятием полисинтетического языка, но ничего не знали о реальных полисинтетических языках; из того, что они читали, было ясно лишь, что это что-то очень сложное. Один из студентов сдал работу со следующим определением: «В “полисинтетическом” языке слова устроены очень сложно. Это значит, что в одном слове содержится более одного значащего элемента, как, например, в английском *cat-s*».

Андерсон определяет сложность как те свойства системы, которые не обязательно следуют из ее природы, из ее неотъемлемых компонентов. Вспомним, что говорили о языке авторы введения: синтаксис и фонология — необходимые компоненты языка, но без морфологии язык мыслим. Андерсон соглашается с этим и идет дальше, разделяя морфологию на морфотактику (организацию морфологического материала в более крупные единицы — слова) и алломорфию (реализацию одного и того же элемента разными способами). Гипотетически можно было бы ожидать, что морфотактика ничем не будет принципиально отличаться от синтаксиса (зачем языку **две** разные системы объединения элементов в более крупные сущности?), а алломорфия — от фонологии. Андерсон показывает, что это не так. Морфотактика не сводится к синтаксису. Например, в языке квакиутль (вакашская семья) действуют жесткие ограничения на порядок слов в предложении: глагол всегда занимает первое место, субъект всегда предшествует объекту, а выраженные прилагательным определения — существительному, являющемуся их вершиной. Помимо того, в квакиутле есть богатая система лексических суффиксов, позволяющих использовать одно сложное слово там, где другие языки обошлись бы двумя (или более) независимыми словами. В сложных словах, вопреки синтаксической закономерности, элемент, который соответствует глаголу, обычно следует за элементом, соответствующим объекту, а определение может следовать за определяемым. Если сложное слово содержит глагол и объект, то объект, таким образом, может предшествовать независимому субъекту. Андерсон также показывает, что алломорфия не сводится к фонологии (приводя пример алломорфии, которая не может быть объяснена в фонологических терминах).

Далее автор ставит перед собой задачу описать «логическое пространство», в которое может укладываться разнообразие морфологической сложности, существующей в языках мира, главные «измерения», по которым в языках возникает затейливая структура слова,

не выводимая очевидным образом из необходимых свойств языка. Измерения эти попадают в две большие категории: свойства системы в целом и свойства отношений между отдельными морфологическими элементами и их экспонентами. Далее Андерсон подробно и с примерами рассматривает несколько измерений в каждой категории (показывая, что именно может являться сложным, то есть не выводимым из базовых свойств системы), я ограничусь простым перечислением. Свойства системы в целом: количество аффиксов в языке, количество аффиксов в слове, принципы объединения аффиксов в слово. Свойства отношений между элементами: сложность реализации отдельных элементов, сложность отношений между словами (имеются в виду члены парадигмы), сложность алломорфии.

Затем Андерсон задается вопросом, откуда берется морфологическая сложность, как она возникает в языке. Один очевидный источник — это грамматикализация. Другой — фонологические изменения (которые, например, могут привести к тому, что фонологическое чередование перестанет быть прозрачным и будет переосмыслено как морфологический показатель). Особо Андерсон отмечает, что сегодняшняя морфология — это не всегда вчерашний синтаксис (вопреки афоризму Талми Гивона), и приводит пример, когда морфологические свойства возникают сами по себе, не из синтаксиса и не из фонологии. В альсайидском жестовом языке (это так называемый деревенский жестовый язык, возникший спонтанно и сравнительно недавно, притом сравнительно хорошо изученный) третье поколение носителей регулярно образует сложные слова, помещая определение перед вершиной: молиться[^]дом ‘мечеть’. В синтаксических же конструкциях порядок слов обратный: вершина предшествует зависимому (педантичности ради следует упомянуть, что в работе [Meir et al. 2010], на которую ссылается Андерсон, этот факт не приводится, но в других работах того же коллектива, например [Aronoff et al. 2010: 142—143], он действительно обсуждается).

В завершение Андерсон отмечает, что, хотя морфология выглядит необязательным компонентом языка, она тем не менее присутствует во многих языках и, более того, нередко возникает в ходе языковых изменений. Вероятно, это не случайно и морфология каким-то (пока неизвестным) образом глубоко укоренена в природе языка и языковых способностей человека. Так что на самом-то деле, заключает Андерсон, наше понимание того, что в морфологических системах сложно, а что нет, пока не слишком глубоко.

На этом кончается **первая часть**.

Третья глава, написанная Эрихом Раундом (Erich Round), называется «Ризоморфемы, мероморфемы и метаморфемы» (Rhizomorphemes, meromorphemes and metamorphemes). Все три термина придуманы автором и опираются на понятие морфомы, тоже не самое широко известное.

Морфома — это категория (или совокупность форм), существующая в морфологической системе, но не изоморфная ни одной категории ни одного иного уровня: семантики, синтаксиса, фонологии. Марк Аронофф [Aronoff 1994: 24—25], который ввел это понятие, приводил в пример английское причастие прошедшего времени, которое служит для выражения перфекта (вместе с глаголом *have*), а также пассива. Пассив и перфект, однако, совершенно разные категории и в семантическом, и в синтаксическом смысле, и ни из чего не следует, что они должны всегда выражаться одной и той же формой. Значит, заключает Аронофф, существует некоторая отдельная функция, которая сопоставляет пассиву и перфекту одно и то же фонологическое выражение. Функция эта не является ни морфосинтаксической, ни морфофонологической, но чисто морфологической. Она существует на особом, морфомном, уровне и называется морфомой. По-русски о морфомах см., например, [Люсина 2013].

Раунд выделяет три типа морфом. Ризоморфемы представляют собой множества корней. Они разбивают словарь языка на классы, элементы которых имеют схожие парадигмы, например, склонения или спряжения. Мероморфемы представляют собой множества формообразовательных операций и разбивают морфологические соответствия (mappings) на классы, элементы которых имеют схожие шаблоны экспонирования. Глава 3 посвящена

в основном мероморфам в языке каядилт (семья пама-ньюнга). Метаморфозы представляют собой множества клеток в парадигме и разделяют типы парадигм (*sic*) на классы, элементы которых имеют схожую реализацию мероморфом. Насколько я могу понять, обсуждавшаяся выше функция, которая сопоставляет одну и ту же форму перфекту и пассиву, — это мероморфома, а множество клеток парадигмы английского глагола, которое мы называем причастием прошедшего времени, — это метаморфома. В целом глава чрезвычайно трудна для чтения.

Показывая, что словоизменение каядилта разумно анализировать с использованием понятия морфом, Раунд отмечает, что морфомы, по крайней мере в данном случае, тоже пример морфологической сложности. Они играют значительную роль в организации морфологической системы, но не повышают ее выразительность. Если принять точку зрения Андерсона, что морфология — это фактически сложность сама по себе (в силу своей необязательности в системе), то морфомы, чисто морфологические категории, не описываемые на иных уровнях, — это сложность в сложности. Подобно авторам двух предыдущих глав, Раунд задается вопросом (на будущее), какую же выгоду может приносить эта дополнительная сложность. В заключение, однако, он уточняет, что, хотя само **существование** морфом — проявление сложности, в **формальной структуре** всех трех типов морфом никакой особенной сложности нет и они крайне мало отличаются от уже известных типов архитектуры грамматических подсистем.

Четвертая глава, написанная Марком Донохью (Mark Donohue), называется «Морфологическая непрозрачность: ссылочные правила в глаголах языка канум» (Morphological opacity: Rules of referral in Kanum verbs). Донохью выделяет четыре различных проявления сложности. Первое — большой размер системы, количество элементов в ней (например, сложная местоименная система может содержать много местоимений). Второе — большое количество измерений системы, большое количество переменных, необходимых для описания системы (в сложной местоименной системе может существовать несколько значений числа, рода и лица). Третье — типологическая редкость элементов системы (сложная местоименная система может содержать маркеры чрезвычайно редко встречающихся различий, например, поколенческой разницы между адресатом и говорящим). Четвертое — непрозрачность: в сложной системе нет прямых соответствий между значениями и их выражением. Продолжая ряд примеров с местоименной системой, Донохью упоминает папуасский язык мериам-мир, где инклюзивное местоимение 1 л. мн. ч. выражается с помощью морфомы 2 л., к которой приобъединен суффикс мн. ч.

Эти четыре проявления не всегда независимы, а вполне могут коррелировать (чем больше в системе измерений, тем больше потенциальное количество элементов, тем больше вероятность возникновения непрозрачных соответствий и т. п.). Донохью концентрируется на непрозрачности, рассматривая глагольное словоизменение в языке канум (Новая Гвинея). В глаголе маркируется субъект, объект и время. Набор различных морфем невелик, и некоторые категории реализуются регулярно. Сложность системы в том, что во многих случаях действуют так называемые ссылочные правила (*rules of referral*): клетку парадигмы занимает показатель из другой клетки. Так, например, в парадигме настоящего времени глагола *ampl* 'насмехаться' маркер объекта 1 л. мн. ч. синкретичен с маркером 2 л. ед. ч., а маркеры 2 и 3 л. мн. ч. синкретичны с маркером 3 л. ед. ч. м. р.

Донохью подробно описывает канумские ссылочные правила и заключает, что, хотя они лексически непредсказуемы, можно установить ряд закономерностей (какие ссылки между клетками в каких случаях возможны). Видимо, именно анализ такого рода Баерман, Браун и Корбет имели в виду, говоря о неочевидной элегантности и сравнимости рассматриваемых в книге систем.

Пятая глава, авторы которой Жан-Пьер Кениг (Jean-Pierre Koenig) и Карин Майкельсон (Karin Michelson), называется «Морфологическая сложность по-онейдски» (Morphological complexity à la Oneida). Онейда (онайда) — ирокезский язык, знаменитый своей системой из 58 местоименных префиксов, которая является обязательной частью глагольного

словоизменения. Так, *luwa-hlo-li-he?* означает 'она, они или кто-либо говорят ему', а *shukwa-hlo-li-he?* — 'он говорит нам'. Дефисы обозначают морфемные границы, *he?* — показатель хабитуалиса, *hlo-li* — корень, *luwa* — префикс со значением 'агнс 3 л. ед. ч. ж. р., неопределенного 3 л., 3 л. дв. ч. или 3 л. мн. ч. совершает действие по отношению к пациенсу 3 л. ед. ч. м. р.', а *shukwa* — 'агнс 3 л. ед. ч. м. р. совершает действие по отношению к пациенсу 1 л. мн. ч.'.

Кениг и Майкельсон подробно описывают онейдскую систему префиксов, обращая особое внимание на такие параметры, как размер, семантическая омонимия, сложность сегментации и «прямота» (directness). Последнее, видимо, эквивалентно прозрачности в терминах Донохью: авторы показывают, что для описания онейдской системы недостаточно «прямых» правил вида «Сопоставь множеству морфосинтаксических значений такую-то форму». Как и для языка канум, необходимо вводить ссылочные правила. Представляется неудачным, что в двух соседних главах практически одно и то же понятие называется по-разному (если между понятиями на самом деле все же есть существенная разница, то она от меня ускользает и в любом случае не обсуждается эксплицитно). На суть изложения это, разумеется, не влияет.

Третий параметр — сложность сегментации — особенно интересен: Кениг и Майкельсон отмечают, что ранее в литературе он для данной системы не обсуждался. Авторы рассматривают, насколько носителю трудно правильно проанализировать данную форму и идентифицировать префиксы. Из-за ряда фонологических явлений эта задача далеко не всегда является тривиальной.

Как и авторы введения, Кениг и Майкельсон подчеркивают, что морфологическая сложность по своей природе не синтагматична, а парадигматична. Если рассматривать сложность не как абстрактный параметр, а как трудность для носителя, то парадигматическая сложность — это трудность **выбора** (формы при порождении, значения при восприятии). Авторы показывают, что носитель онейдского языка, чтобы использовать правильный префикс, должен сделать выбор между 326 алломорфами 58 морфов (здесь было бы естественно ожидать термин «морфема», но авторы выбирают «морф», «чтобы не ограничивать себя морфемным подходом к словоизменению», с. 82). По субъективным оценкам тех, кто учит онейдский во взрослом возрасте, это чрезвычайно трудно (поскольку детей, осваивающих онейдский как первый язык, уже нет, то неизвестно, насколько это было бы трудно для них). Но только ли в размере системы дело? Словарь любого языка во много раз больше, но его освоение обычно не ассоциируется с такими же трудностями. Кениг и Майкельсон предлагают следующие гипотетические факторы, объясняющие, почему выбор онейдского префикса — трудная задача.

Во-первых, выбор префикса обязателен. Описывая ситуацию, говорящий всегда должен употребить один и только один префикс. Во-вторых, выбор зависит от достаточно общих свойств ситуации, которые не обязательно актуальны для собеседников. Говорящий должен определить, сколько в ситуации участников, одушевленные ли они, определенные ли, какой род и лицо следует употребить и т. п. Авторы утверждают, что говорящему может быть необходимо сделать до девяти бинарных решений, чтобы выбрать нужный морф. В-третьих, в системе присутствуют случаи нейтрализации. С одной стороны, они облегчают задачу говорящего, потому что уменьшают количество необходимых решений, с другой — усложняют, потому что нейтрализация несистематична: нет общих простых правил, которые хорошо описывали бы, какие клетки парадигмы синкретичны. Кроме того, нейтрализация усложняет задачу слушающего. Отвлекаясь от конкретного онейдского случая, надо сказать, что эта двоякая роль нейтрализации (синкретизма, ссылочных правил) в языковой сложности — большая проблема. С одной стороны, уменьшение числа экспонентов является упрощением, с другой, меньшая прозрачность правил, к которому оно нередко приводит, является усложнением. Какая английская парадигма сложнее: *cat* — *cats* или *moose* — *moose*? К сожалению, в книге этот вопрос систематически не рассматривается, хотя анализ синкретизма занимает центральное место во многих главах. Возвращаясь

к главе Кенига и Майкельсон: в-четвертых, выбрав морф, говорящий должен затем выбрать алломорф, что тоже не вполне тривиальная задача (и к тому же усложняет задачу сегментации для слушающего).

Можно, конечно, усомниться в том, что в реальной речи выбор и анализ морфов устроены именно так, как предполагает приведенное выше описание. Например, абстрактивный подход к морфологии (см. восьмую главу), вероятно, предполагает иную модель. Авторы и не утверждают, что их модель точно соответствует когнитивной реальности. Это попытка приблизительно оценить трудности, создаваемые для носителей языка сложными парадигмами, — достаточно умозрительная, но весьма интересная.

Шестая глава, написанная Мариной Чумакиной (Marina Chumakina) и Гревилем Корбетом (Greville Corbett), называется «Классно-числовое маркирование в арчинском языке: сложное в малом» (Gender-number marking in Archi: Small is complex). Арчинский язык знаменит, например, своими гигантскими глагольными парадигмами. Чумакина и Корбет обращаются, однако, к парадигме небольшой: множеству маркеров числа и согласовательного класса. Хотя маркеров немного, сложность проявляется в том, насколько непредсказуемо устроено согласование.

Любопытно, что авторы, судя по всему, считают, что в фокусе этой главы тоже находится парадигматическая сложность — но, поскольку речь идет о согласовании, я бы классифицировал наблюдаемые явления хотя бы отчасти как синтагматическую сложность.

Сложность заключается в непредсказуемости того, что и как будет согласовываться. Практически все части речи могут получать классно-числовые маркеры, но не все лексемы одной и той же части речи ведут себя одинаково. Более того, маркеры иногда могут быть инфиксами, а иногда — префиксами, что представляется типологической редкостью. Чумакина и Корбет подробно анализируют систему, выявляя зависимость наличия и вида маркирования от частеречных, морфологических, морфофонологических и фонологических факторов. Тем не менее даже после установления ряда закономерностей остается существенное количество исключений, которые приходится специфицировать на лексическом уровне.

На этом кончается **вторая часть**.

Седьмая глава, написанная Грегори Стампом (Gregory Stump) и Рафаэлем Финкелем (Raphael Finkel), называется «Сравнение различных способов представления словоизменительных систем и некоторые его следствия для вычисления морфологической сложности» (Contrasting modes of representation for inflectional systems: Some implications for computing morphological complexity). Основная мысль главы заключается в том, что результат измерения морфологической сложности, основанного на некоем формальном представлении словоизменительной системы, в очень большой степени зависит от выбранного способа представления. Способов же таких может быть много, и выбор самого подходящего далеко не всегда очевиден. Это чрезвычайно ценное наблюдение, важность которого невозможно преувеличить. Я совершенно согласен с авторами, что эта зависимость результата измерения от способа представления не всегда осознается исследователями в достаточной степени.

Стамп и Финкель иллюстрируют свое утверждение следующим образом. Они представляют словоизменительные системы с помощью **карт** (они используют термин «plat», что в переводе с американского английского означает примерно ‘кадастровая карта’) — таблиц, в которых колонки представляют множества морфосинтаксических свойств (ММС), строки — словоизменительные классы, а в клетках находятся экспоненты соответствующего ММС для соответствующего класса. Разумеется, составляя карту для данного языка, исследователь должен принять ряд решений (выбрать ММС, выбрать классы, выбрать способ представления экспонентов), которые влияют на итоговую картину.

Для 40 английских глаголов Стамп и Финкель составляют две карты: в одной экспоненты представлены в фонемной транскрипции, в другой — в морфофонемной транскрипции с добавлением границ морфов. Первая карта содержит формы, непосредственно доступные слуховому восприятию, поэтому авторы называют ее ориентированной на слушателя

(ОНС); вторая же отражает представления носителя языка о структуре слова, поэтому ее называют ориентированной на говорящего (ОНГ). См. различия двух карт на примере двух глаголов в таблице 1.

Существенно, что для данных глаголов можно, разумеется, составить не две, а сколько угодно карт, которые с точки зрения ориентации на говорящего или слушающего, скорее всего, составят континуум между двумя довольно радикальными картами, приведенными в главе 7.

Таблица 1

Две карты двух английских глаголов

Лексема	{инф}	{ЗедНастИзьяв}	{прош}	{настПрч}	{прошПрч}
Карта, ориентированная на слушающего					
BAND	ænd	ændz	ændəd	ændiŋ	ændəd
BUILD	ɪld	ɪldz	ɪlt	ɪldiŋ	ɪlt
Карта, ориентированная на говорящего					
BAND	ænd	ænd-z	ænd-d	ænd-iŋ	ænd-d
BUILD	ɪld	ɪld-z	ɪld-t	ɪld-iŋ	ɪld-t

Стамп и Финкель рассматривают несколько мер сложности: условную энтропию («неуверенность» в предсказании экспонента для ММС X на основании знания экспонента для ММС Y), свойства множеств основных частей (principal parts; множеств клеток, из которых можно вывести содержание остальных клеток), предсказуемость (predictability) словоизменительного класса, предсказуемость клетки и предсказательную силу (predictiveness) клетки. Они показывают, что карта ОНГ получает более низкие значения энтропии, более высокие значения предсказуемости и предсказательной силы и имеет меньше основных частей. Таким образом, карта ОНГ оказывается по всем предложенным мерам проще карты ОНС, главным образом потому, что в ней содержится грамматическая информация, отсутствующая в ОНС. Из этого, разумеется, не следует, что ОНГ лучше, а только лишь, что карты **разные**.

Очевидно, что можно критиковать выбранные авторами способы представления, предложенные меры, а также сам «картографический» подход. Неясно, например, как представить в картах случаи, которые Андерсон во второй главе приводит в качестве примеров сложности парадигмы: вариативность (в одной клетке более одной формы) или дефективность (в клетке нет формы, как в первом лице настоящего времени единственного числа глагола *победить*). Из главы не очевидно, насколько высока надежность вычислений (анализ карт и вычисления мер производились автоматически). Вероятно, ответы на эти вопросы можно получить из других работ авторов (например, [Stump, Finkel 2013]); программа-анализатор, кроме того, доступна свободно. Для основного пафоса главы, однако, все это не очень существенно, а с ним, повторюсь, невозможно не согласиться.

Восьмая глава называется «Вычислительная сложность абстрактивной морфологии» (Computational complexity of abstractive morphology, о переводе слова *abstractive* см. ниже) и написана Вито Пиррелли (Vito Pirrelli), Марчелло Ферро (Marcello Ferro) и Клаудией Марци (Claudia Marzi). Начинают авторы с противопоставления конструктивного (constructive) и абстрактивного (abstractive) подходов к морфологии. Мне неизвестно, существует ли устоявшийся перевод этих терминов на русский, поэтому я принимаю самое простое решение и заимствую их. Ярлыки «конструктивный» и «абстрактивный» вводятся в работе [Blevins 2006], но идея этого противопоставления гораздо старше.

При конструктивном подходе основными морфологическими строительными блоками являются корни и аффиксы. После того как носители выучивают, как выглядят части слов, они более не хранят в памяти полные формы, каждый раз порождая или анализируя их

заново. При абстрактивном подходе основной единицей являются именно полные формы, они хранятся в памяти, а освоение морфологии — это освоение отношений между этими формами, которое приводит к выделению «сублексических единиц».

Авторы описывают новый подвид нейронных сетей, способных обучаться морфологии. Глава 8 исключительно трудна для читателей, не имеющих серьезной подготовки в соответствующей области. Основная идея авторов заключается в том, что радикально абстрактивный подход, на который опираются их нейронные сети, дает им ряд преимуществ, которые и демонстрируются в главе. В заключение Пиррелли, Ферро и Марци отмечают, что нейронаука уже довольно много знает о «низкоуровневых» свойствах мозга, а лингвистика — о «высокоуровневых» свойствах языковых структур. Посередине — пропасть, которая затрудняет взаимодействие двух дисциплин. Компьютерные модели вроде рассмотренных в главе 8 — одна из возможностей заполнить эту «среднеуровневую» пропасть.

Девятая глава называется «Схемы синкретизма и парадигматическая сложность на примере древне- и среднеиндийского склонения» (Patterns of syncretism and paradigm complexity: The case of Old and Middle Indic declension). Ее автор Паоло Милицья (Paolo Milizia) рассматривает проявления гипотетического принципа компенсаторности, предложенного в свое время Вигго Брёндалем [Brøndal 1940]: в маркированной категории обычно меньше различий, чем в противопоставленных ей немаркированных. Классический пример — синкретизм падежей в дв. ч. имен в древних индоевропейских языках. Одно из возможных объяснений принципа компенсаторности — стремление избежать чрезмерно большого числа маркированных значений [Cairns 1986: 18]. Так, например, латинский аблатив совпадает во мн. ч. с дативом, потому что значение и падежа, и числа маркировано. Милицья задается резонным вопросом, в каком смысле маркированные значения могут быть аддитивны и в каком месте морфологической системы происходит их сложение. Он предполагает, что значения сами по себе складываться не могут, но тем менее маркированность можно считать аддитивной, если интерпретировать ее как частотность (по мнению автора, этот подход совместим, например, с критикой понятия маркированности Хаспельматом [Haspelmath 2006]). Тогда объяснение синкретизма состоит в том, что в морфологических системах есть тенденция избегать распределений частотности экспонентов, слишком удаленных от равномерного, поэтому клетки двух или более редких морфосинтаксических свойств могут получать общий экспонент. Остается, правда, непонятным, **зачем** языку стремиться к равномерным распределениям, какие преимущества из этого следуют.

Милицья вводит две меры, позволяющие квантифицировать степень удаленности реального распределения от равномерного. Одна — разность энтропий этих двух распределений: $D = \log_2 N - H(X)$, где X — рассматриваемое распределение, H — энтропия, N — количество экспонентов в равномерном распределении (его энтропия максимальна и равна логарифму N). Эта мера эквивалентна расстоянию Кульбака — Лейблера между распределениями. Другая мера — избыточность: $R = 1 - H(X)/\log_2 N$. При совпадении наблюдаемого распределения с равномерным обе меры равны нулю. Это любопытный взгляд: я, например, привык воспринимать энтропию именно как меру сложности (чем выше энтропия, тем выше неопределенность, непредсказуемость распределения, тем труднее оно для освоения и употребления), из чего следует, что изменения, приводящие к оптимизации системы (с точки зрения человека, изучающего язык), обычно будут приводить к уменьшению энтропии. Милицья же исходит из того, что система оптимальна при максимальном значении энтропии.

Далее Милицья на примере адъективного склонения на *-a/ā* в древнеиндийском языке и языке пали (как синхронных наблюдений, так и диахронических сопоставлений) стремится показать, что синкретизм, возможно, возникает не случайно, а так, чтобы эффективно уменьшать D и R . Рассматриваемые парадигмы, по мнению автора, представляют собой контрпример к принципу компенсаторности в формулировке Брёндаля (потому что для женского рода во мн. ч. различается больше косвенных падежей, чем в ед. ч.), но не в новой формулировке в терминах частотности и теоретико-информационных свойств. Милицья также приводит примеры из латыни, старославянского и русского.

Милиция отмечает, что предполагаемая тенденция к равномерному распределению лишь один из факторов, влияющих на структуру парадигмы. Это совершенно разумное замечание, но оно приводит к существенному недостатку главы 9. Примеры, вписывающиеся в гипотезу автора, признаются подтверждающими ее, а не вписывающиеся объясняются влиянием других факторов: семантики, функциональных соображений, тенденции высокочастотных экспонентов быть короче, особого статуса женского рода в праиндоевропейском и т. п. Опять же, вполне может быть, что так и есть, но при таком подходе очень легко впасть в нефальсифицируемость. Чтобы убедительно обосновать гипотезу автора, нужен или более строгий количественный подход, показывающий, что влияние стремления к равномерности статистически достоверно (оно при этом может быть слабым), или строгий принцип обращения с контрпримерами и альтернативными объяснениями, а еще лучше — и то и другое. Учитывая, что и сама гипотеза представляется несколько надуманной, глава 9 выглядит довольно неубедительно. С другой стороны, эта глава в значительной степени представляет собой краткий пересказ книги автора [Milizia 2013], вероятно, там можно найти некоторые ответы на эти сомнения.

Любопытно, что меры Милиции — единственные количественные меры в книге, которые можно отнести к классу корпусных (corpus-based), а не грамматических (grammar-based): они основаны не только на абстрактном представлении парадигмы, но и на реальных частотных распределениях. К корпусным данным эксплицитно обращаются также авторы пятой и шестой глав, но эти главы не позиционируются как количественные.

Учитывая ориентацию книги на парадигматическую сложность и общий довольно абстрактный (не путать с абстрактным!) подход к морфологии, доминирование грамматических мер неудивительно. В общем и целом я считаю более перспективными корпусные меры, но очевидно, что у них есть ряд своих неустраняемых недостатков, от которых свободны грамматические меры. Кроме того, корпусные меры, если в корпусе есть хоть какая-нибудь разметка, тоже опираются на грамматические модели, поэтому разработку «абстрактного грамматического понимания» языковой сложности можно только приветствовать.

Десятая глава, написанная Себастьяном Банком (Sebastian Bank) и Йохеном Троммером (Jochen Trommer), называется «Обучение и сложность нулевого выражения» (Learning and the complexity of Ø-marking) и посвящена автоматической сегментации слова на морфемы. Авторы фокусируются не на практических приложениях алгоритмов сегментации, а на их способности моделировать усвоение сегментации человеком и на теоретических выводах, которые из этого можно сделать (это не их эксплицитное заявление, а моя интерпретация главы).

Банк и Троммер замечают, что задача **стемминга** (выделения основ) решена уже довольно хорошо, но вот методы детального анализа аффиксальной строки (той части слова, которая остается после стемминга), то есть разделения ее на аффиксы, пока разработаны мало. В главе 10 авторы абстрагируются от несегментной и неконкатенативной морфологии, не рассматривают инфексию и случаи, когда порядок аффиксов друг относительно друга может варьировать (как в арчинском).

Авторы вводят четыре класса алгоритмов, способных обучаться сегментации. Все алгоритмы перебирают все возможные анализы (анализ состоит из разбиения аффиксальной строки на аффиксы и сопоставления аффиксов элементам значения) и выбирают тот, который наилучшим образом соответствует заданным критериям. Различаются алгоритмы тем, какие разбиения они способны перебирать. Для их описания нам понадобятся понятия свободного и связанного употребления аффикса. При свободном употреблении аффикс совпадает с аффиксальной строкой; при связанном употребляется как подстрока в аффиксальной строке. Например, употребление *-st* во втором лице единственного числа настоящего времени в таблице 2 — свободное, прошедшего времени — связанное, а у *-s* вообще нет свободных употреблений.

Алгоритм нулевого класса неспособен делить аффиксальную строку на отдельные аффиксы. Результат его работы для парадигмы *glauben* совпадал бы с таблицей 2. Алгоритм первого класса способен выделять в аффиксальной строке те аффиксы, у которых

встречаются свободные употребления. Для алгоритма второго класса ограничение слабее: он способен разбивать строку S на $S_1 + S_2$, если хотя бы одна из строк S_1 или S_2 имеет свободное употребление (S_1 и S_2 могут затем подвергаться дальнейшему разбиению). Алгоритм третьего класса не имеет никаких ограничений и способен перебирать все мыслимые разбиения.

Разумеется, чем больше номер класса, тем мощнее алгоритм: он способен делать все то же самое, что и алгоритмы классов с меньшими номерами, и еще кое-что. Но вместе с номером растет (причем взрывообразно) и размер пространства поиска — количество возможных разбиений и сопоставлений, которое алгоритму приходится перебирать. Поэтому авторы предполагают, что парадигмы, для успешного анализа которых необходимы алгоритмы классов 2 и 3, будут в языках мира встречаться редко.

Таблица 2

Результат стемминга немецкого глагола *glauben* ‘верить’¹

	Настоящее время		Прошедшее время	
	ед. ч.	мн. ч.	ед. ч.	мн. ч.
1	<i>glaub-e</i>	<i>glaub-(e)n</i>	<i>glaub-te</i>	<i>glaub-ten</i>
2	<i>glaub-st</i>	<i>glaub-t</i>	<i>glaub-test</i>	<i>glaub-tet</i>
3	<i>glaub-t</i>	<i>glaub-(e)n</i>	<i>glaub-te</i>	<i>glaub-ten</i>

Далее Банк и Троммер задают критерии, по которым алгоритмы будут выбирать наилучший анализ. Очевидно, что далеко не всегда языковая реальность позволяет установить одно-однозначные соответствия между аффиксами и значениями, и столь же очевидно, что необходимые отклонения от одно-однозначного идеала будут выглядеть по-разному в зависимости от выбора анализа. Авторы показывают, что эти отклонения можно описать в привычных компьютерным лингвистам терминах «точности» (precision) и «полноты» (recall) и, более того, что морфологи фактически обычно пользуются эквивалентными критериями, только не столь формализованными и квантифицированными. Ранжирование критериев может быть устроено по-разному, в зависимости от того, на какую морфологическую теорию опираться (авторы стремятся к максимальной точности, что порой может приводить к не максимальной полноте).

Таблица 3

Сегментация *glauben* алгоритмом первого класса

	Настоящее время		Прошедшее время	
	ед. ч.	мн. ч.	ед. ч.	мн. ч.
1	<i>glaub-e</i>	<i>glaub-n</i>	<i>glaub-te</i>	<i>glaub-te-n</i>
2	<i>glaub-st</i>	<i>glaub-t</i>	<i>glaub-te-st</i>	<i>glaub-te-t</i>
3	<i>glaub-t</i>	<i>glaub-n</i>	<i>glaub-te</i>	<i>glaub-te-n</i>

Затем авторы анализируют глагольные парадигмы нескольких языков с помощью алгоритмов разных классов. Для успешного анализа парадигмы *glauben*, например, достаточно алгоритма первого класса, см. разбиение в таблице 3 (в 1 и 3 л. мн. ч. наст. вр. отсутствует *-e*-, так как авторы практически абстрагируются от фонологических явлений, обучая алгоритмы с помощью глубинных представлений (сноска 3 на с. 187)) и соответствия аффиксов значениям:

te [+прош] *st* [+2 +ед] *e* [+1 +ед –прош]
n [–2 +мн] *t₁* [+2 +мн] *t₂* [+3 +ед –прош]

¹ Основы отделены от аффиксальных строк дефисами.

Это, отмечают авторы, примерно соответствует традиционному анализу немецкой глагольной парадигмы. Если к тому же глаголу применить алгоритм второго класса, то получится другой анализ, который тоже обсуждается в литературе [Müller 2005]. Отличие заключается в том, что этот алгоритм способен вычленив аффикс *-s-*, то есть *-st* во 2 л. ед. ч. обоих времен в таблице 3 распадается на *-s-t*. Тогда из множества соответствий исчезает *st*, но появляется *s* [+2 +ед], а значение *t_i* изменяется на [+2]. Авторы оговаривают, что они не убеждены в том, что анализ Мюллера лучше традиционного, но тот факт, что алгоритмы разных классов выдают анализы, хорошо соответствующие тем, которые выдвигают разные лингвисты, им кажется примечательным.

Бывают, однако, языки, с которыми алгоритм первого класса не справляется: такова, например, глагольная парадигма суахили. Авторы находят только один случай, где необходим алгоритм третьего класса: глагольная парадигма ленакельского языка (австронезийская семья).

Ключевое лингвистическое наблюдение авторов состоит в том, что сложность сегментации парадигмы (номер необходимого класса) падает, если в ней встречается нулевое выражение. Так, например, в немецкой парадигме 1 и 3 л. ед. ч. прош. вр. выражаются нулем, что дает свободное употребление *te*, что, в свою очередь, упрощает дальнейшую сегментацию. Таким образом, хотя нулевое выражение в литературе часто рассматривается как источник сложности (см. ссылки на с. 187), оно может рассматриваться и как фактор, способствующий сегментации. Возможно, его столь широкое распространение в языках мира не случайно.

Банк и Троммер выдвигают гипотезы о том, что морфологические системы адаптируются к выучиваемости, избегая сложности третьего класса (полный перебор) и стремясь к сложности первого класса. Они проводят пилотную проверку на небольшой типологической выборке из 20 языков, рассматривая глагольные парадигмы (согласование с субъектом и маркеры времени, вида (аспекта) и модальности (наклонения), или ТАМ). Поскольку непосредственное вычисление сложности — задача исключительно трудоемкая, авторы рассматривают присутствие нулевого выражения. Так, второго класса может быть достаточно только в том случае, если хотя бы один маркер ТАМ или хотя бы один маркер согласования — нуль. Это выполняется для 19 языков (единственное исключение — уже упоминавшийся ленакельский). Первого класса может быть достаточно только в том случае, если есть хотя бы один нулевой маркер ТАМ и хотя бы один нулевой маркер согласования. Это выполняется для 11 языков.

На этом кончается **третья часть**, а с ней и книга.

Интересной неожиданностью для меня стала позиция редакторов-составителей (которую явно разделяет как минимум часть авторов) по отношению к вопросу: «Зачем изучать языковую сложность?». Вообще говоря, мыслимы разные ответы. Самый простой — зачем, это не нужно (я не встречал его эксплицитно в письменном виде, но слышал устно). Другой ответ — чтобы использовать параметр «сложность» для предсказания иных свойств языка. Например, для практических целей: оценивать, насколько трудно выучить данный иностранный язык. Успешных межязыковых сравнений такого рода я не знаю, но сравнивать сложность текстов в пределах одного языка вполне возможно (о сравнении текстов на русском языке и о роли морфологии в оценке сложности см., например, [Reynolds 2016]).

В обзоре [Бердичевский 2012] я даю следующий ответ: понимание и измерение сложности интересны не столько сами по себе, сколько как часть некоей объяснительной теории. Там же я рассматриваю одну из самых влиятельных таких теорий (точнее, группу родственных теорий): сложность языка зависит от социальных параметров, в первую очередь от интенсивности контактов с носителями других языков. Чем выше в популяции доля тех, кто осваивает язык как неродной, тем выше вероятность, что морфологическая сложность будет падать.

Достоинство этой теории в том, что она позволяет отвечать на вопросы, которые редакторы сборника оставили открытыми, в первую очередь — почему языки имеют разную сложность. На вопрос Андерсона о том, почему же морфологическая сложность вообще

возникает в языке, существует два ответа (оба весьма гипотетические). Одна гипотеза гласит, что сложность — это груз нефункциональных «украшений», накапливаемых тысячелетиями в результате более или менее случайных процессов, если социальные условия это позволяют [McWhorter 2001; Dahl 2004]. Другая — что сложная морфология «вредна» для взрослых, но является «полезной» для детей, способствуя усвоению ими языка [Lupyan, Dale 2010; Ellis 2008].

Редакторы сборника тоже, разумеется, знакомы с соответствующей литературой (это видно из обсуждения на с. 7). Тем не менее они сознательно воздерживаются от последнего шага и не встраивают сложность в объяснительные теории. Этот отказ от типологического и социолингвистического подхода к сложности роднит книгу с другим недавним сборником — «Измерение грамматической сложности» [Newmeyer, Preston 2014].

Тем не менее для меня остается не вполне ясным, зачем отдельно изучать сложность, если не пытаться ни объяснить ее распределение в языках мира, ни использовать сложность как фактор для предсказания каких-то иных свойств. Можно возразить, что сборник представляет шаг в этом направлении: прежде чем что-либо **объяснять** про сложность, следует сначала описать, осмыслить и измерить то разнообразие сложности, которое наблюдается в языках мира.

Выходя за рамки рецензии конкретного сборника, отмечу, что вопросами о том, почему же сложность языков различается и в каких случаях она может исчезать, Суррейская группа задается в своем следующем проекте «Утрата словоизменения» (Loss of inflection).

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Бердичевский 2012 — Бердичевский А. Языковая сложность (language complexity) // Вопросы языкознания. 2012. №5. С. 101—124. [Berdičevskij A. Language complexity. *Voprosy jazykoznanija*. 2012. No. 5. Pp. 101—124.]
- Вахтин 2014 — Вахтин Н. Б. [Рец. на:] P. Trudgill. Sociolinguistic typology: Social determinants of linguistic complexity // Антропологический форум. 2014. № 22. С. 301—309. [Vakhtin N. B. [Review of:] P. Trudgill. Sociolinguistic typology: Social determinants of linguistic complexity. *Antropologičeskij forum*. 2014. No. 22. Pp. 301—309.]
- Люсина 2013 — Люсина В. С. [Рец. на:] M. Maiden, J. Ch. Smith, M. Goldbach, M. Hinzelin (eds.). Morphological autonomy: Perspectives from Romance inflectional morphology // Вопросы языкознания. 2013. № 2. С. 135—141. [L'usina V. S. [Review of:] M. Maiden, J. Ch. Smith, M. Goldbach, M. Hinzelin (eds.). Morphological autonomy: Perspectives from Romance inflectional morphology. *Voprosy jazykoznanija*. 2013. No. 2. Pp. 135—141.]
- Aronoff 1994 — Aronoff M. *Morphology by itself: Stems and inflectional classes*. Cambridge (MA): MIT Press, 1994.
- Aronoff et al. 2010 — Aronoff M., Meir I., Padden C., Sandler W. The roots of linguistic organization in a new language. *The Emergence of protolanguage: Holophrasis vs compositionality*. Amsterdam: John Benjamins, 2010. Pp. 133—152.
- Blevins 2006 — Blevins J. Word-based morphology. *Journal of Linguistics*. 2006. No. 42. Pp. 531—573.
- Brøndal 1940 — Brøndal V. Compensation et variation, deux principes de linguistique générale. *Scientia*. 1940. No. 68. Pp. 101—109.
- Cairns 1986 — Cairns C. Word structure, markedness and applied linguistics. *Markedness. Proceedings of the Twelfth Annual Linguistics Symposium of the University of Wisconsin-Milwaukee*. New York: Plenum, 1986. Pp. 13—38.
- Dahl 2004 — Dahl Ö. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins, 2004.
- Ellis 2008 — Ellis N. The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal*. 2008. Vol. 92. No. 2. Pp. 232—249.
- Haspelmath 2006 — Haspelmath M. Against markedness (and what to replace it with). *Journal of Linguistics*. 2006. No. 42. Pp. 25—70.
- Lupyan, Dale 2010 — Lupyan G., Dale R. Language structure is partly determined by social structure. *PLoS ONE*. 2010. Vol. 5. No. 1. e8559.

-
- McWhorter 2001 — McWhorter J. The world's simplest grammars are creole grammars. *Linguistic Typology*. 2001. Vol. 5. No. 2—3. Pp. 125—166.
- Meir et al. 2010 — Meir I., Sandler W., Padden C., Aronoff M. Emerging sign languages. *Oxford handbook of deaf studies, language and education*. Vol. 2. Marschark M., Spencer P. (eds.). New York: Oxford Univ. Press, 2010. Pp. 267—280.
- Milizia 2013 — Milizia P. *L'equilibrio nella codifica morfologica*. Roma: Carocci, 2013.
- Müller 2005 — Müller G. *Subanalyse verbaler Flexionsmarker*. Leipzig: Universität Leipzig, 2005.
- Newmeyer, Preston 2014 — Newmeyer F., Preston L. (eds.). *Measuring grammatical complexity*. Oxford: Oxford Univ. Press, 2014.
- Reynolds 2016 — Reynolds R. Insights from Russian second language readability classification: Complexity-dependent training requirements, and feature evaluation of multiple categories. *Proceedings of the 11th Workshop on innovative use of NLP for building educational applications*. Association for Computational Linguistics, 2016. Pp. 289—300.
- Stump, Finkel 2013 — Stump G., Finkel R. *Morphological typology: From word to paradigm*. Cambridge: Cambridge Univ. Press, 2013.