

ОТНОШЕНИЕ СВОБОДНОГО ПОРЯДКА СЛОВ И МОДЕЛИ ИНВЕРСИИ¹

FREE WORD ORDER AND MODELLING OF INVERSION

А.В.Циммерлинг (meinmat@yahoo.com)
МГОПУ им. М.А.Шолохова

Принимается, что язык L имеет свободный порядок, если относительное расположение двух любых произвольно взятых категорий X и Y может быть обращено. Производную последовательность можно вывести из исходной путем отображения последней целиком, но можно получить ее и за счет передвижения одного элемента, приписав ему свойство подвижности.

Фундаментальной чертой порядка слов в естественных языках является его предсказуемость: на основании системы правил и запретов, свойственных конкретному языку, можно предсказать, какие линейные последовательности элементов предложения будут грамматически правильными или оптимальными, а какие — запрещенными или нежелательными. Предсказуемость порядка элементов в предложении до некоторой степени сближает синтаксис естественных языков с синтаксисом формальных и является предпосылкой для создания автоматизированного парсера, т.е. устройства, членившего синтаксическую структуру на более мелкие составляющие и выстраивающего эти составляющие в правильной линейной последовательности. Создание парсеров для естественных языков — не только прикладная, но и фундаментальная теоретическая задача: оно позволяет выявить механизмы, распознающие и порождающие правильные синтаксические структуры и, тем самым, успешно изучать любой язык как естественнонаучный объект.

В большинстве работ по типологии порядка слов при определении отношения «фиксированного» и «свободного порядка» синтаксических элементов исходят из запретительных механизмов. При этом встает проблема выбора исходной или инвертированной последовательностей, поскольку все языки мира хотя бы в некоторых случаях допускают более одного грамматичного порядка элементов предложения. Отношения свободного порядка при таком подходе определяется негативно, как отсутствие жестких запретов на линейные альтернативы. Более перспективным нам кажется подход, при котором во главу угла ставятся разрешительные механизмы, а отношение «свободного порядка слов» будет определено через понятие инверсии, а не наоборот. Варьирование порядка слов уместно интерпретировать как характеристику класса естественных языков. Но в языках, допускающих большое количество линейных перестановок элементов предложения, формализовать интуицию носителей данных языков, отделяющую разрешенные последовательности от нежелательных, трудно, в связи с чем возникает ложное представление о полном отсутствии ограничений. Отражением его и является метафора «свободный порядок слов». В научной литературе класс языков с т.н. свободным порядком обычно выделяется по двум разным критериям:

- невозможности выделить какой-то один доминирующий порядок развертывания категорий предложения (обычно определяемых на базе небольшого числа предположительно универсальных категорий, например, Глагол – Подлежащее – Дополнение vs Подлежащее – Глагол – Дополнение);
- Отсутствие или незначительно малое количество формальных запретов на перестановку элементов предложения (ср., например, правило, по которому спрягаемый глагол в немецком языке должен занимать второе место в независимом повествовательном предложении).

Парадоксальным образом, первый признак не обязательно предполагает наличие второго, и наоборот. Так, например, в хеттском и латинском языках наиболее частотным и стилистически немаркированным был порядок SOV (Глагол – Подлежащее – Дополнение), однако в этих древних индоевропейских языках были возможны и стилистически маркированные порядки вида VSO (с начальным глаголом) и SV (с начальным подлежащим). Ср.

(1) šallanun= war=an kuit ammuk (StBoT 24iv 11)

я-возвеличил=мол=его ибо Я

“Ибо я (сам) мол его возвеличил”

(2) # Aparrus=ma=kan išparzašta (КВо 2.5. ii 35-36)

Апарру=но=ЧАСТИЦА он-ускользнул

“<X. Осадил Апарру и убил его людей.> Но Апарру сумел ускользнуть”

¹ Данная работа выполнена при финансовой поддержке гранта РГНФ № 06-04-00203а

Примеры (1) и (2) подтверждают, что в хеттском языке не было формального запрета, блокировавшего перемещение предиката и подлежащего в менее типичные для них позиции. Тем не менее, хеттский и латинский относятся к классу языков, для которых гипотеза о доминирующем порядке слов остается в силе. Противоположный случай засвидетельствован в древнескандинавских языках. Здесь имеется один полностью формализованный запрет — на перемещение спрягаемой части глагольного сказуемого дальше второй позиции от начала простого повествовательного предложения, но относительное расположение прочих категорий предложения могло быть любым, что доказано в (Циммерлинг 2002: 316-353).

В данной статье класс языков со свободным порядком слов обсуждается в двух аспектах: 1) в терминах стратегий кодирования, определяющих порядок развертывания предложения в данном языке; 2) в терминах формальных ограничений, задающих систему порядка слов. Принимается следующее оперативное определение отношения свободного порядка:

- Язык L имеет свободный порядок элементов предложения, если относительное расположение двух любых произвольно взятых категорий предложения X и Y может быть обращено: $[X + Y] \Rightarrow [Y + X]$

Принципиальное отличие избранного подхода от большинства практикуемых состоит в том, что класс языков со свободным порядком элементов предложения определяется с опорой на понятие линейной перестановки — относительной инверсии. Кроме нее, выделяется абсолютная инверсия, т.е. изменение позиции категории предложения по отношению к началу синтаксической структуры (ср. «начальная» позиция, «конечная позиция», «2,3.. n-ная позиция от начала» и т.п.), и обсуждаются методы разметки синтаксической структуры.

Упомянутые выше древнескандинавские языки (древнеисландский, древненорвежский, древнешведский, древнедатский и древнегютнийский) попадают в класс языков со свободным порядком, точно так же, как современный русский. В то же время русский язык, в отличие от древнескандинавских, не входит в класс языков с отсчетом позиций от начала предложения. В последний класс попадают все германские языки, кроме современного английского и вымершего готского, языки с т.н. Ваккернагеля и клитиками, занимающими второе место в предложениях, а также языки с запретами на дистантное размещение сентенциальных клитик и глагольных форм. К последней группе относятся современные болгарский и македонские языки, а также ряд центрально-филиппинских языков (тагальский, бикол, себуано). Ср. в тагальском языке:

- (3a) kahapon *ka ba nila* nakita?
 Вчера 2SgSbj ли 3Pl.DO быть-увиденным
 “Вчера ли они тебя видели?”
- (3б) Nakita *ka ba nila* (kahapon)?
 быть-увиденным 2SgSbj ли 3Pl.DO (вчера)
 “Они тебя видели (вчера)?”

Механизм отсчета позиций от начала предложения может действовать лишь при наличии формальных запретов на размещение одной или нескольких категорий предложения. Так, в тагальском и болгарском языках группу клитик (ср. таг. *ka ba nila*) нельзя сдвинуть со второй позиции, а глагольные формы, при наличии клитик в составе предложения, должны занимать либо первую, либо третью позицию: #XP—Clitic —Vf ~ #Vf — Clitic, #*XP — ...—Clitic, #*XP — Clitic ...—Vf, #*Clitic—Vf, *Vf —...Clitic².

К языкам со свободным порядком слов, удовлетворяющих данному выше определению, применяется следующая гипотеза:

- Если язык L имеет свободный порядок слов, порядок полноударных элементов предложения в нем определяется стратегиями кодирования коммуникативных отношений, т.е. совокупностью механизмов, создающих блоки, объединенные не формально-синтаксической общностью (члены предложения, группы), а общей коммуникативной маркировкой (имеющими статус темы, ремы, контрастной темы, контрастной ремы).

К языкам со значительным количеством неполноударных слов (клитик), важных для построения предложения (ср. безударные местоимения и связки в славянских и романских языках), применяется дополнительная гипотеза:

- Позиции клитик и словоформ-хозяев, к которым они присоединяются, задаются стратегией просодического кодирования, разбивающего предложение на фонетические слова (тактовые группы), объединенные одним ударением.

Без перебора большого числа языков мира неясно, как соотносятся языки с коммуникативным и просодическим кодированием: включает ли один из этих классов другой, или же они пересекаются. Важно установить также, совместимо ли просодическое кодирование с фиксированным порядком слов.

Анализ и обработка материала связаны с двумя группами проблем. Первая касается выбора исходной линейной последовательности:

- Если в языке L со свободным порядком слов предложение, состоящее из n элементов ($n \geq 2$), может быть развернуто более чем одним способом, требуется определить, какая линейная последовательность является исходной или нейтральной, и объяснить почему.

² Символ ‘XP’ читается ‘произвольная синтаксическая категория’, ‘Vf’ — ‘финитная форма глагола’, символ ‘#’ обозначает границу предложения.

Решение этой проблемы требует применения комплексных методов, в частности, изучения условий ветвления (branching condition, направление развертывания группы от ядра к зависимому элементу, либо наоборот), статистического анализа (частотность альтернирующих порядков), и анализа в терминах контекстной обусловленности/коммуникативной нагрузки порядка слов.

Вторая группа проблем связана с тем, как *вывести производную* линейную последовательность из исходной. Решение этой задачи требует сопоставления естественных и формальных языков. Имеются две теоретические альтернативы:

- 1) Производная линейная последовательность В {b + a + c} получается из исходной последовательности А {a + b + c} путем отображения последней целиком: перемещения элементов a, b, c по отдельности не рассматриваются.
- 2) Производная линейная последовательность В {b + a + c} получается за счет передвижения какого-то одного элемента (например, b), в то время как прочие элементы своих позиций не меняют. Один из вариантов этой концепции предполагает анализ в терминах *следов* перемещаемых элементов и *пустых синтаксических позиций*, ср. возможную запись: {a + b + c} ⇒ {b + a + t_b + c}.

Второй подход обладает большей объяснительной силой, но является и более обязывающим, так как критерии выбора перемещаемого элемента не всегда очевидны. Кроме того, в группах из 4 и более элементов возможны перестановки двух и более элементов сразу, что не редкость в языках мира. Ср. в русском языке:

(4) Речи (a), которые (b) никогда (c) раньше (d) не слышали (e) в Лондоне (f) из уст мэра (g) {abcdefg} ⇒ речи (a), которые (b) раньше (d) в Лондоне (f) из уст мэра (h) никогда (c) не слышали (e) {abdgfce}.

Такие последовательности нельзя получить друг из друга за один шаг.

Апробируется гипотеза о том, что:

- В языках со свободным порядком слов перемещаемый элемент (словоформа, группа или блок слов) может быть успешно вычленен, и б) перемещение элемента никогда не является бессмысленной операцией, она всегда преследует ту или иную коммуникативную цель.

Так, при анализе русской пары предложений *Вася (a) приходит (b) домой (c) ⇒ приходит (b) Вася (a) домой (c)* целесообразно считать, что второе предложение получается из первого за счет перемещения глагола в начало предложения. Планом содержания этой операции является расщепление рематической составляющей {приходит домой} с помещением уже известной слушателю темы {Вася} внутрь последней.

В настоящее время общепризнано, что все множество правильно построенных предложений не может быть порождено при помощи т.н. базового компонента синтаксиса, т.е. при помощи Контекстно-Свободных грамматик (КС-грамматик), поскольку при этом игнорируются многие инвертированные линейные порядки, разрешенные в языках мира, ср. (Гладкий, Мельчук 1969: 95, 115). Одной из задач ближайшего будущего является формализация подклассов Контекстно-Зависимых правил (КЗ-правил). Наиболее ценны те модели парсинга, которые можно с минимальной настройкой использовать для разных языков со свободным порядком слов.

Один из типов перемещения элементов, объяснимый только в терминах КЗ-правил, заслуживает особо пристального внимания. Речь идет о варьировании позиций сентенциальных и глагольных клитик, в зависимости от линейного контекста. Ср. пример (5) в древнерусском языке:

(5) #того (1)=ся (2) каю (3), «я каюсь в том». Клитика *ся*, по стандартному для данного языка правилу, стоит в препозиции глаголу, но отходит на шаг вправо, если ей предшествует более одной ударной словоформы, и переходит в постпозицию глаголу, ср. пример (6) в том же древнерусском языке:

(6) #Того (1) всего (2) каю(3)=ся (4) «я каюсь во всем том» (Зализняк 2004: 60).

Ценность подобных примеров для общего синтаксиса в том, что они подтверждают наличие перемещений, заданных КЗ-правилами и действующих *слева направо*. Между тем, в Минималистской Программе Н.Хомского и связанных с ней доктринах генеративной грамматики санкционируются лишь перемещения *справа налево*, поднимающие элементы в вышестоящие позиции в дереве предложения: это влечет за собой контринтуитивные объяснения и допущения, не мотивированные свойствами языковых объектов. Однако некоторые КЗ-правила линеаризации не удастся адекватно объяснить и при классическом структуралистском подходе. Так, А.А.Зализняк трактует случаи отхода сентенциальных клитик вправо дальше второй позиции в языках с законом Ваккернагеля как несовпадение фонетического и синтаксического начала предложения, т.н. Правило Барьера: «начальная часть может быть как бы отчленена... правила о расположении энклитик действуют после этого так, как если бы отчлененной части предложения вообще не было (Зализняк 1989: 286; ср. Зализняк 2004: 54-56). Аналогичное КЗ-правило, под другим названием, было постулировано А. Хальперном. Однако в ряде языков отход клитик вправо меняет позиции других элементов, и гипотеза о том, что первая группа в предложении не занимает синтаксической позиции, не годится. Так, в словацком языке возвратные местоимения (sa, si) и местоимения дат.п. в possessивной функции обычно занимают вторую позицию; место глагола при этом несущественно.

(7)

XP	Clitic	(YP)	VP
1	2	3	4
aby	si	□lovek	[predstavil _srazu]

что	<i>себе=Refl</i>	человек	[представил __ сразу]
-----	------------------	---------	-----------------------

«чтобы человек мог сразу себе представить»

Если же начальная группа оказывается длинной и синтаксически сложной, клитики отодвигаются на шаг вправо, освобождая вторую позицию для глагольных форм:

(8)

XP	Vf	Clitic	YP
1	2	3	4
Vodi□ autobusu Bus driver	zap<lil put fire	Si to-oneself =Refl	Cigaretu Cigarette

«Водитель автобуса зажег (для себя) сигарету»

Мы видим, что КЗ-правило влияет на позицию сразу двух категорий предложений, при этом осуществляются две разных операции: сдвиг клитик вправо и перемещение финитного глагола.

Одной из задач ближайшего будущего является содержательное и формальное описание неканонических перемещений, не охватываемых существующими версиями генеративной грамматики. Это позволило бы внедрить более гибкие синтаксические модели в типологические описания языков мира.

Последняя группа проблем касается размежевания компетенции синтаксиса и просодии, и различия между устными и письменными текстами. Известно, что в языках со свободным порядком слов одна и та же линейная последовательность слов может быть коммуникативно неоднозначна и произноситься с разными фразовыми акцентами, ср. рус. Иван поедет в ДЕРЕВНЮ, ИВАН поедет в деревню, Иван ПОЕДЕТ в деревню и т.п. Поэтому для полноценного описания отношения свободного порядка необходимо учитывать не только чисто линейные, но и *линейно-акцентные преобразования* (термин введен Е.В.Падучевой), меняющие позицию и/или акцентуацию элемента предложения. Факты, накопленные к настоящему времени, побуждают выдвинуть следующие гипотезы:

- Линейно-акцентные преобразования служат планом выражения изменения тема-рематических отношений в классе языков со свободным порядком слов.
- Множество линейно-акцентных преобразований в языках мира ограничено небольшим числом, что в ряде случаев позволяет вывести коммуникативный статус элементов из их линейной комбинаторики и реконструировать коммуникативно значимые преобразования в репрезентативном корпусе письменных текстов в древних и новых языках со свободным порядком слов.

По-видимому, ни одна из существующих доктрин не представляет оптимальной базы для проверки данных гипотез. В современных версиях генеративной грамматики, в частности, в Минималистской Программе Н.Хомского, формализация отношения свободного порядка сопряжена с трудностями. Оба распространенных подхода хомскианцев к языкам типа современного русского — стандартная концепция т.н. перемещения лексических категорий и альтернативная ей концепция т.н. Скрэмблинга (букв. «перемешивания») — имеют ряд недостатков. При первом подходе выделяются исходная и конечная позиции перемещаемого элемента, но а priori принимается постулат о наличии фиксированного порядка формально-синтаксических категорий во всех языках. При втором подходе многообразие позиций элемента (например, альтернатива в языке порядков Подлежащее – Дополнение – Глагол ~ Дополнение – Подлежащее – Глагол) интерпретируется как неграмматическое явление, и синтаксис линейных отношений предстает в виде набора бессмысленных формальных операций. Бесспорным достижением генеративистов является формализация в 1970-е гг. феномена инверсии в терминах перемещения, пустой синтаксической позиции и следа перемещенной категории. Но в существующих версиях грамматики Хомского и близких к ним доктринах санкционируются лишь перемещения, действующие *справа налево*, что неприемлемо. Прорыв в области изучения свободного порядка слов был достигнут в пионерских работах 1930-1950-х гг. о коммуникативной перспективе высказывания (В.Матезиус, Ф.Данеш, П.Адамец). Продолжением этой традиции следует считать работы И.Е.Ковтуновой и Е.В.Падучевой; последней введено фундаментально важное понятие линейно-акцентных преобразований, инвариантом которых является лексико-синтаксическая структура предложения с заданным числом словоформ. Но эти концепции до сих пор не получили типологического применения. То же приходится констатировать применительно к Модели «Смысл ⇒ Текст» И.А.Мельчука: предложенная Мельчуком многоступенчатая процедура применения разных КЗ-правил для вывода альтернирующих линейных порядков громоздка и мало подходит для парсера. В части работ по типологическому синтаксису принимаются априорные допущения о порядке слов. Другая часть авторов эти допущения отвергает, подчеркивая статистическую природу тенденций, действующих в сфере порядка слов (ср. школу Дж.Хокинса). В 1990-2000-е гг. бурно развивается типология клитик (А.Цвикки, А.Хальперн, Ст.Андерсон, А.Айхенвальд): большинство зарубежных работ в этой области можно считать пограничными между типологией и общей грамматикой. Такие работы еще в большей степени, нежели исследования в области передвижения полноударных элементов, перегружены лингвистической догматикой, создающей проблемы, вместо того, чтобы решать их: так, перемещение клитик вправо либо отрицается вообще, либо признается «просодическим» явлением, якобы не имеющим отношения к синтаксису. Подобные ссылки на просодию или морфологию, когда инвентарь собственно синтаксических объяснений ограничен постулатами генеративной

доктрины, стали общим местом. В последние десятилетия на западе получили распространение неортодоксальные генеративные концепции (GPSG, HPSG, Tree-adjoining Grammar), которые ориентированы на преодоление разрыва между жесткой доктриной Хомского и эмпирикой языков с большим количеством линейных перестановок. Достоинством этих грамматик является то, что при их помощи удастся охватить некоторые классы высказываний, порождаемых КЗ-правилами (т.н. mildly-context-sensitive rules), и распознать некоторые непроективные синтаксические структуры с пересечением составляющих.

Список литературы

1. *N.Chomsky*. A Minimalist Program for Linguistic Theory // The view from building 20. /Hale, K. S.L.Keyser (eds). Cambridge, Mass. MIT Press. 1993.
2. Clitics in the languages of Europe. Eurotyp 20-5. /Ed. by H. van Riemsdijk. Mouton de Gruyter. Berlin- New York. 1999. 1020 S.
3. *R. Fiengo*. On Trace Theory // Linguistic Inquiry 8, 1977. 35-62.
4. *G.Gazdar, E.Klein, G.Pullum, I. Sag*. Generalized Phrase Structure Grammar. Harvard University Press. Cambridge, Mass, 1985.
5. *J.H.Greenberg*. Some universals of grammar with particular reference to the order of the meaningful elements // Universals of language, Cambridge, Mass. 1963, p. 73-113.
6. *S. Luraghi*. The pragmatics of Verb Initial Sentences in Some Ancient Indo-European Languages // Word Order in Discourse, Downing, Noonan 1995, 335-386.
7. *А.В.Гладкий, И.А.Мельчук*. Элементы математической лингвистики. М.,Наука, 1969.
8. *А.А. Зализняк*. «Слово о Полку Игореве»: взгляд лингвиста. М., Языки Славянской Культуры, 2004.
9. *Е.В.Падучева*. Коммуникативная структура предложения и понятие коммуникативной парадигмы // НТИ, Сер. 2. 1984. N 10.
10. *А.А.Перекрестенко*: Об автоматическом синтаксическом анализе в некоторых классах контекстно-зависимых языков //Московский лингвистический журнал, том 6, №2. Москва, 2003.
11. *А.А.Перекрестенко*. Разработка парсера и модуля унификации для синтаксического процессора // Системы и средства информатики, М.: Наука, выпуск 14, 2005.
12. *А.В. Циммерлинг*. Типологический синтаксис скандинавских языков. М., «Языки славянской культуры», 2002. 896 стр.
13. *А.В.Циммерлинг*. Порядок слов и синтаксические позиции. // Труды международного семинара "Диалог 98" по компьютерной лингвистике и её приложениям /А.С. Нариньяни (ред.). Казань, 1999.
14. *А.В. Циммерлинг*. Порядок слов в славянских, германских и романских языках. Славяно-германские исследования. Спб, Алетейя, 2006 (в печати).