

НЕКОТОРЫЕ ВОПРОСЫ МОРФОЛОГИЧЕСКОГО АНАЛИЗА ПРИ МАШИННОМ ПЕРЕВОДЕ С ВЕНГЕРСКОГО ЯЗЫКА НА РУССКИЙ

(Критические замечания)

Ш. КОНИ

Одни специалисты по машинному переводу руководствуются высказыванием Шлегеля «Lisez, lisez; jetez vos grammaires au feu»¹ и отодвигают грамматический анализ текста на второй план, значит используют грамматический разбор для улучшения, а не для обоснования перевода², другим, особенно советским исследователям представляется чрезвычайно важным сформулировать правила грамматического разбора, так как метод пословного перевода может быть применен только к языкам, структура которых имеет большое сходство³.

Р. Х. Риченс и А. Д. Будс, которые первыми дали пример автоматического анализа венгерского текста, привели такое предложение: Az apró* bogyó*jú fajták, úgy termés*mennyiség*ben, mint száraz* anyaghozam*ban, felülmúl*ják a nagy*bogyó*jú fajták*at⁴. Звездочки отделяют первую основу от второй (в сложных словах) или основу от окончания. (На границе существительных *anyag* и *hozam* не поставлена звездочка, но это, по-видимому, опечатка, судя по их английским эквивалентам *matter* и *yield*. Звездочка эта явно ошибочно переместилась в абсолютный конец слова *száraz*).

Здесь интересно наблюдать, что по этому разбору существительное *fajták* не распадается на основу *fajta* (или *fajta-*) и на показатель множественного числа *-k*, а после переводящего слова *variety* поставлена помета *m* (множественное число). А дальше слово *fajtákat* переводится следующим образом: *variety m a* (винительный падеж множественного числа). Мы можем констатировать, что составителям анализирующих правил такого рода хорошо известны были сложные слова, но с такими простыми существительными, в которых после основы стоят морфемы (больше одной), они ничего не могли делать.

¹ Цит. по сборнику статей *Machine Translation of Languages. Fourteen essays*, New York — London, 1957, стр. 24.

² Ср. *Linguistic and Engineering Studies in the Automatic Translation of Scientific Russian into English; Technical Report Phase II*. Seattle, 1960.

³ Ср., например: Д. Ю. Панов. Автоматический перевод, М., 1956. — О. С. Кулагина и И. А. Мельчук. Машинный перевод с французского языка на русский. — ВЯ, 1956, № 5, стр. 111—121.

⁴ R. H. Richens — A. D. Booth. Some methods of mechanized translation; *Machine Translation of Languages. 14 essays*. 1955, стр. 37.

Несравнимо подробнее разработал правила машинного анализа венгерских слов И. А. Мельчук.⁵ Его работе мы дали оценку в двух статьях, стараясь дополнить и исправить полученные И. А. Мельчуком результаты.⁶ В третьей статье мы занимаемся анализом венгерских существительных⁷.

В последнее время, как нам кажется, в связи с готовящимся обратным словарем венгерского языка была выдвинута мысль начинать морфологический анализ с последней буквы слова⁸. Такой подход нам представляется неприменимым для целей машинного перевода. Чтобы определить один только падеж венгерского *существительного*, мы должны были бы сперва выбрать из текста «традиционным способом» все существительные, а после этого передать их для обработки машине.

Так как предложенный метод не включает в себя пользования словарем основ, только таблицей падежных окончаний, для него «найти падежное окончание» значит найти максимальное по числу букв окончание в таблице. Целью упрощающееся в слове (поиск начинается в обратном порядке, с последней буквы.) Чтобы машина не ошиблась в определении падежа, для нее нужно составить специальный список венгерских существительных, концы которых в начальной форме совпадают с каким-либо падежным окончанием (прилепом). Список должен содержать такие слова, как *isten, Balaton, akarát, kanál, szoba* и т. п. и за неимением особых правил и такие существительные, как *asztal*, чтобы машина не проанализировала в перечисленных словах прилепы *-en, -on, -at, -nál, -ba, -tal*. (В связи с последним окончанием имеется в виду вариант прилепа *-val*.) Кроме «общезвестных» вариантов этого прилепа, по орфографическим соображениям нужно было бы ввести и варианты *-al, -el*, чтобы машина могла определить падеж (социатив), например, существительных *tekercs, modell*. Мы не говорим теперь о том, что такого рода омонимы, как *falat* «кусочек» и «стену» (не считая третье значение в глаголах: «заставить жрать») окажутся по предложенному анализу всегда существительными в винительном падеже (если этого слова нет в особом списке) или же наоборот, существительными в именительном падеже (если упомянутый список содержит

⁵ И. А. Мельчук. О машинном переводе с венгерского языка на русский. Проблемы кибернетики, вып. I, М., 1958, стр. 222—264. В венгерском переводе: I. A. Melcsuk, A magyarról oroszra történő gépi fordítás szabályainak ismertetése: Matematikai nyelvészet és gépi fordítás a Szovjet-unióban, Budapest, 1964. Időszerű Műszaki Dokumentációs Kérdések, №6, стр. 147—214.

⁶ Kónyi Sándor. A magyar—orosz gépi fordítás néhány problémája; Általános nyelvészeti tanulmányok II. Budapest, 1964, стр. 197—204. — III. Кони. Рецензия на статью И. А. Мельчука «О машинном переводе с венгерского языка на русский». Computational Linguistics III. Budapest, 1964, стр. 226—243.

⁷ Kónyi Sándor. A magyar főnevek elemzése: Általános nyelvészeti tanulmányok III. Budapest. Сборник сдан в печать.

⁸ Ср. Magyar Nyelvőr, 1965. №1. стр. 143.

в себе данное слово.) Так как для такого анализа интересен только падеж, все равно, возьмем ли слово *áramot* «ток» или *áramat* «мое шило и т. п.», результат будет одинаковым: винительный падеж, и мы можем только жалеть, что о слове *áramat* машина больше информации не дала.

Однако нам представляется, что составить вышеупомянутый особый список невозможно, просто из-за большого количества отглагольных существительных, оканчивающихся на *-at*, *-et* (*nézet*, *váгат*, *váгат*, *furat*, *vállalat*, *képviselet*, *folýamat*). Далее нам совсем неясно, как проанализировать такие слова, как *eben*, *kara* или *kára* (сравни со словоформами *éвben* и *fára*, где анализ по предложенному методу возможен.) Или же нужно применить другой специальный список?

А если взять не одни существительные, но все части речи, или же попробовать разобрать по этому методу все «подходящие» слова (значит существительные) целых предложений «без посторонней помощи», то мы увидим, что такой анализ столкнется с непреодолимыми трудностями. Станут проблематичными не только существительные *ének*, *asztal*, но и глаголы *ötenek*, *vígasztal* и т. д.

В ходе морфологического анализа венгерских текстов нужно определить не только падеж существительных (или вообще имен), значит нужно рассматривать не только последнюю морфему имени, но все морфемы всех изменяемых частей речи, ведь в венгерском языке иногда даже наречия могут принимать падежные окончания: *alul*, *alulra*, *alulról*⁹. Мы видели, что, по мнению Риченса и Будса, после основы венгерского слова может стоять только одна морфема, но И. А. Мельчук уже правильно подчеркивал, что в силу агглютинативного строя венгерского языка «окончание» (то, что остается после отделения основы или двух основ сложного слова) в венгерском языке не едино (как в английском, французском или русском); оно может состоять из нескольких суффиксов, каждый из которых имеет одно значение¹⁰. И. А. Мельчук не определяет число морфем, которые могут стоять с «машинной» точки зрения после основы, однако судя по его таблицам и примерам *módosulásaival*, *nehézségeire*, можно полагать, что, по его мнению, таких морфем три. По нашим представлениям позади основы может стоять шесть морфем, например, *legfelhasználhatóbbakéiban* «в принадлежащих наиболее используемым», «в принадлежащих тем, кого можно лучше всего использовать», не считая префикса превосходной степени *leg-* и отделяемой приставки *fel-*, существующих основе. Осложненный притяжательный признак обладателя *-éi* не разбит на элементы *-é* и *-i*; таким образом получилось шесть морфем, а не семь. (Думается, что И. А. Мельчук неправомерно разбивает сложные личные притяжательные окончания *-ai*, *-ei* и т. п. на две морфемы.)

Ознакомившись с трудностями автоматического анализа венгерских слов, читатель легко может себе представить, что полный морфологический разбор

⁹ См. *Telegdi Zsigmond. Elméletek a szófajok természetéről* (Kritikai megjegyzések): *Nyelvtani tanulmányok*. Szerk. Sulán Béla. Budapest, 1961, стр. 45.

¹⁰ Проблемы кибернетики, вып. I, М., 1958, стр. 230.

не удастся, если анализ начнется с конца слова. Мы думаем не только о таких длинных словах, как *legfelhasználhatóbbakéiban*, но и о таких, как *hajójárat* «пароходная линия», *gyermekkoromban* «(я) в детстве», *tibennünk* «в нас». Если взять, например, последнее слово, то по таблице глагольных суффиксов найдем информацию к *-nünk*: неопределенная форма, 1 лицо множественного числа; *-ben* — (по другой таблице) окончание инессива и т. д.; сложное слово *gyermekkoromban* может быть ошибочно переведено на русский язык так: «в детской саже» и т. д.

Эти примеры подчеркивают еще раз, что анализ должен начинаться с первой буквы слова, ведь (первая) основа более или менее определяет дальнейший ход анализа, в то время как последний остаток «не в силах» этого сделать.

Возьмем к примеру словоформу *szótára* «его словарь». Если нет словаря основ, машина может ошибочно найти, например окончание сублатива *-ra*, (остается бессмысленное *szótá-*), или может разбить словоформу на компоненты *szót* «слово (вин. п.)» и *ára* «его цена и т. п.». (В современном венгерском языке не имеется таких сложных слов, состоящих из двух существительных, где первый компонент принимал бы прилеп винительного падежа *-t*.)

Мы видим, что единственным правильным подходом при анализе может быть следующий подход: Сначала искать слово в словаре основ, а только потом «остаток». Таким образом машина найдет в словаре сложные существительные *szótár* или *gyermekkor* (ср. с нашим предыдущим примером *gyermekkoromban*), которых нет необходимости больше разбирать, имея в виду цели непосредственного перевода с венгерского языка на русский, так как русские эквиваленты упомянутых слов — простые существительные «словарь» и «детство». «Остатки» *-a* и *-omban* уже легко можно найти в таблице окончаний, если — как это вообще бывает — при основах есть помета, что они являются основами существительных.

В связи с задачами непосредственного перевода с венгерского языка на русский возникает вопрос, стоит ли разбирать сложные венгерские слова, или же воспринимать их как одно целое. Мы думаем не только о том, что в русском языке сложных существительных гораздо меньше, чем в венгерском, но и о том, что на основании «составных частей» венгерских существительных не всегда удастся «построить» их русские эквиваленты.

Мы постарались дать ответ на возникший вопрос на основе рассмотрения венгерских сложных слов и их русских соответствий. Мы собрали слова разных семантических значений. (Слова эти взяты из двухтомного словаря Л. Хадровича и Л. Гальди¹¹.)

¹¹ Magyar—oroszc szótár. (Szerk: Hadrovics L., Gáldi L.) Budapest, 1964.

Венг. сложные слова		Русские эквиваленты		%
1-й комп.	2-й комп.	1-й компонент	2-й компонент	
vese	любой	почеч(ный)	переводящее слово 2-го компонента	27,2
		переводящее слово 2-го компонента	псчек	36,3
		переводящее слово 2-го компонента	почками	4,5
		Простое слово		27,2
		почев(ой)	переводящее слово 2-го компонента	9,0
		переводящее слово 2-го компонента	в почки	4,5
újság	любой	газетн(ый)	переводящее слово 2-го компонента	38,2
		переводящее слово 2-го компонента	газет	44,1
		Производное слово		5,8
		Простое слово		14,7
		переводящее слово 2-го компонента	в газетах	5,8
		газето-	переводящее слово 2-го компонента	2,9

Венг. сложные слова		Русские эквиваленты		%	
1-ый комп.	2-й комп.	1-й компонент	2-й компонент		
vas	любой	железн(ый)	переводящее слово 2-го компонента	57,0	
		стальн(ой)	переводящее слово 2-го компонента	3,0	
		железо-	переводящее слово 2-го компонента	4,4	
		переводящее слово 2-го компонента	железа	6,6	
		Простое слово			21,4
		переводящее слово 2-го компонента	чугуна	0,7	
		переводящее слово 2-го компонента	по железу	0,7	
		переводящее слово 2-го компонента	железняка	0,7	
		металлургическ(ий)	переводящее слово 2-го компонента	5,9	
		железист(ый)	переводящее слово 2-го компонента	0,7	
		графитов(ый)	переводящее слово 2-го компонента	0,7	
		предельн(ый)	переводящее слово 2-го компонента	0,7	
		железно-	переводящее слово 2-го компонента	0,7	
		неприкосновенн(ый)	переводящее слово 2-го компонента	0,7	
		искусственн(ый)	переводящее слово 2-го компонента	0,7	

(Примечание к таблице. Проценты оказываются всегда выше ста, потому что венгерские сложные слова имеют в русском языке иногда больше соответствий.)

Следовательно, на вопрос «стоит ли разбирать венгерские сложные слова», мы даем положительный ответ, так как в большинстве случаев существуют простые способы передачи первой основы сложных слов. Важнейшие из этих способов: образование прилагательного с помощью суффиксов и постановка переводящего слова в родительном падеже. (Притом русские эквиваленты вторых основ могут браться просто из словаря.) Какую из этих возможностей выбрать — это семантический вопрос.