



ЕЗИКЪТ И НЕГОВАТА МНОГОЗНАЧНОСТ



1. ЕЗИКЪТ И НЕГОВАТА МНОГОЗНАЧНОСТ

“Това е кучето, което разтревожи котката, убила плъха, който изяде малца, намиращ се в къщата, построена от Джак.”

Майката гъска, *Къщата, която построи Джак*

“Това е малцът, който плахът, убит от котката, притеснена от кучето, изяде.”

Виктор Х. Ингве



1. ЕЗИКЪТ И НЕГОВАТА МНОГОЗНАЧНОСТ

В следващите редове ще представим набор от формални механизми за моделиране на човешките езици. Включващи крайните автомати, моделите на Марков, преобразувателите, фонетични правила за стилизиране, и безконтекстните граматики.

Преди това ще обсъдим основните формални прилики и разлики между тези модели и как да преценим кой от тях е подходящ за моделиране на определени аспекти от естествения език.



1. ЕЗИКЪТ И НЕГОВАТА МНОГОЗНАЧНОСТ

Ще започнем с представянето на йерархията на Чомски, теоретичен инструмент, позволяващ ни сравняване на изразителната “сила” и “многозначност” на различните формални механизми като автоматите и безконтекстните граматики.



1.1. ЙЕРАРХИЯТА НА ЧОМСКИ

Как са свързани автоматите, безконтекстните граматики и фонетичните правила?

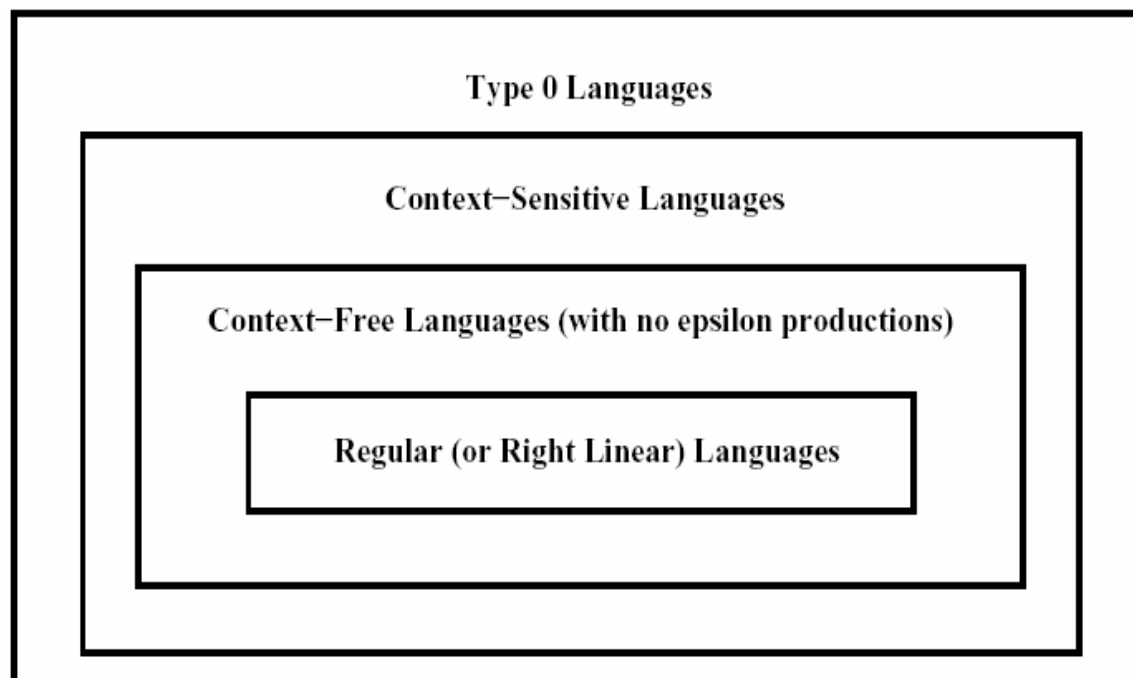
Чрез начинът, по който описват формалния език, виждан във всяка поредица низове от крайна азбука. Но за различните граматики, с които пишем, всеки способ притежава различна генеративна **мощ**.



1.1. ЙЕРАРХИЯТА НА ЧОМСКИ

Възможно е да построите йерархия на граматиките, където съвкупността от езиците, описани от граматики с по-големи възможности включват множество от езици, определящи се от “по-слаби” граматики. Съществуват много такива възможни йерархии; една от най-често използваните в компютърната лингвистика е Йерархията на Чомски.

1.1. ЙЕРАРХИЯТА НА ЧОМСКИ



Фигура 1. Йерархия на Чомски

Където **Type 0 Languages** -> Нулев тип-Езици от общ (нулев) тип- при тях не се налагат никакви ограничения, **Context-Sensitive Languages**-> Контекстно-зависими езици

(Контекстен /първи/ тип), **Context-Free Languages (with no epsilon productions)** (Безконтекстен /втори/ тип) -> Контекстно-независими езици (без изнасяне на епсилон) и **Regular (or Right Linear) Languages** -> Автоматни езици.



1.1. ЙЕРАРХИЯТА НА ЧОМСКИ

Следващата таблица (Фигура 2) показва четирите типа от граматики в йерархията на Чомски, дефинирани чрез ограниченията върху формата, която правилата трябва да вземат.

Тип	Общо наименование	Правило	Лингвистичен пример
0.	Еквивалент на Тюринг	$\alpha \rightarrow \beta, \text{ s.t. } \alpha \neq \epsilon$	Фонологични правила
1.	Контекстна	$\alpha A \beta \rightarrow \alpha \gamma \beta, \text{ s.t. } \gamma \neq \epsilon$	Фразеологична граматика
2.	Безконтекстна	$A \rightarrow \gamma$	Автомати с крайни състояния
3.	Автоматна	$A \rightarrow xB \text{ or } A \rightarrow x$	



1.1. ЙЕРАРХИЯТА НА ЧОМСКИ

За безконтекстните граматики се споменава в Глава 9. Техните правила разрешават всеки единичен нетерминал да се преработи като някакъв низ от терминали и нетерминали. Нетерминалът може също да се сведе до ε , макар че като възможност това не е споменато досега.

Автоматните граматики са еквивалентни на регулярните изрази. Тоест, даденият автоматен език може да се характеризира с всеки регулярен израз от типа, обсъден по-рано в Глава 2, или чрез автоматни граматики. Които граматики от своя страна могат да бъдат дясно и ляво-контекстни (с ляв и десен контекст).



1.1. ЙЕРАРХИЯТА НА ЧОМСКИ

Нататък ще обсъдим само дясно-контекстните граматикки. За пример да вземем следната автоматна (с десен контекст) граматика:

$$S \rightarrow aA$$

$$S \rightarrow bB$$

$$A \rightarrow aS$$

$$B \rightarrow bbS$$

$$S \rightarrow \varepsilon$$



1.1. ЙЕРАРХИЯТА НА ЧОМСКИ

Пример за пораждаване на езика:

$$S \Rightarrow aA \Rightarrow aaS \Rightarrow aabB \Rightarrow aabbbS \Rightarrow aabbbaA \Rightarrow aabbbaaS \Rightarrow aabbbaa$$

Тук забелязваме, че при всяко разширяване на S се получава aaS или $bbbS$; по този начин четеца би трябвало да разпознае сам, че този език отговаря на регулярния израз: $(aa \cup bb)^*$.

Ще представим доказателство, че езикът е автоматен тогава и само тогава, когато е генериран чрез автоматен език; това за пръв път е доказано от Чомски (Chomsky) и Милър (Miller), 1958 г.



1.2. КАКВО ДА НАПРАВИМ, АКО ЕЗИКЪТ НЕ Е АВТОМАТЕН?

Специално за формалните езици съществуват методи за разрешаването на посочените въпроси. Тоест, за конкретен език можем да укажем дали е представим чрез регулярен израз или изисква безконтекстна граматика, и т.н. Така, че ако искате да разберете дали някоя част от естествен език, (фонетичната на Английския или да кажем морфологичната на Турския) е представима чрез известен клас граматика, е нужно да намерим формален език, който моделира съответното явление и изчислява кой точно клас граматика е подходящ за този формален език.



1.2. КАКВО ДА НАПРАВИМ, АКО ЕЗИКЪТ НЕ Е АВТОМАТЕН?

Например, ако Английският беше автоматен, бихме писали регулярни изрази и използвали ефикасните автомати за превод на правилата. Ако Английският е безконтекстен, бихме използвали контекстно-независими правила и алгоритъма на Ърли (Earley), за да парсираме изреченията, и т.н.

Би било добре да знаем къде даден език “запазва” своята сложност; дали фонологичната система на езика е по-проста, в сравнение със синтактичната, или дали определена вид морфологична система е наследствено по-проста от друга.



1.2. РАЗШИРЯВАЩАТА ЛЕМА

Най-обичайният начин да докажеш, че един език е автоматен е да построиш реален регулярен израз за него. Постъпвайки така, може да разчитаме на факта, че автоматните езици са затворени над операциите обединение, конкатенация (слепване), звездата на Клийни, допълнение и сечение.

Някой път искаме да докажем, че даден език не е автоматен. Извънредно полезен инструмент за това е **Разширяващата лема**. Има два извода, произхождащи от тази Лема (нашето описание на Помощната лема е очертано от Луиз /Lewis/ и Пападимитроу /Papadimitriou/ (1981 г.), и Хопкрофт /Hopcroft/ и Улман /Ullman/ (1979 г.).



1.2. РАЗШИРЯВАЩАТА ЛЕМА

Нека разгледаме даден език L и кореспондиращия му детерминиран автомат с крайни състояния M , който има N състояния. Да кажем, че входният низ също е с дължина N . Машината стартира в състояние q_0 ; след преминаване на един символ, минава в състояние q_1 ; след N символа, ще бъде в състояние q_N .

С други думи, низ с дължина N ще мине през $N + 1$ състояния (от q_0 до q_N). Но в машината има само N състояния. Това означава, че поне 2 от тях по продължение на приемника (да ги наречем q_i и q_j) трябва да са еквивалентни.

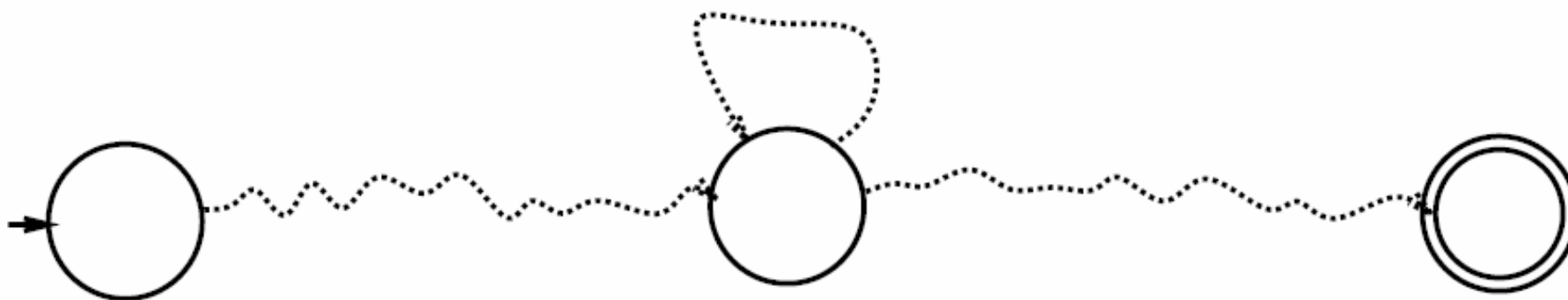


1.2. РАЗШИРЯВАЩАТА ЛЕМА

Машината приема конкатенацията на тези три низа от символи, например xuz . Но ако приема xuz , то би трябвало да приеме и xz ! Това е, защото машината, може просто да прескочи цикъла в процес на xz . Даже още повече- тя би могла да го заобикаля всеки път; така също да приема $xuuz$, $xuuuz$, $xuuuuz$, и т. н. Фактически, трябва да приеме всеки низ $xu^n z$ за $n \geq 0$.

1.2. РАЗШИРЯВАЩАТА ЛЕМА

Фигура 3. Машина с N състояния, приемаща низа xuz от N символа





1.2. РАЗШИРЯВАЩАТА ЛЕМА

Разширяващата Лема. Нека L да бъде автоматен език, разпознаван от детерминиран краен автомат. Тогава съществуват низовете x , y и z , такива че $y \neq \varepsilon$ и $xy^n z \in L$, за $n \geq 0$.

Тази Лема гласи, че ако езикът е автоматен, то съществува някакъв низ y , който може да “разширява/ нараства”, влагайки се по подходящ начин. Но това не означава непременно, че той ще е автоматен. Има и неавтоматни езици, при които имаме разширение. Така, че Лемата не се използва за доказателство, че един език е автоматен. А по-скоро да укажем обратното, показвайки че в някои език не може да се вложи такъв низ подходящо.



1.2. РАЗШИРЯВАЩАТА ЛЕМА

Нека да използваме Лемата, за да покажем, че езикът $a^n b^n$ (тоест езикът съдържа равен брой а-та и b-та, които се редуват един след друг: $ababab\dots$), не е автоматен. Трябва да изведем, че всеки възможен низ s , който посочим, не може да се раздели на три части x , y и z , такива че y да е разширение. Даден произволен низ s от $a^n b^n$, може да се разпадне по три начина, но ще видим, че който и от тях да изберем, не ще намерим подходящия низ y , който да бъде разширен.



1.2. РАЗШИРЯВАЩАТА ЛЕМА

И така, не съществува низ в езика $a^n b^n$, който да се раздели на x , y и z , по начин, по който y да може да се разшири, следователно $a^n b^n$ не е автоматен. Но затова пък е безконтекстен език.

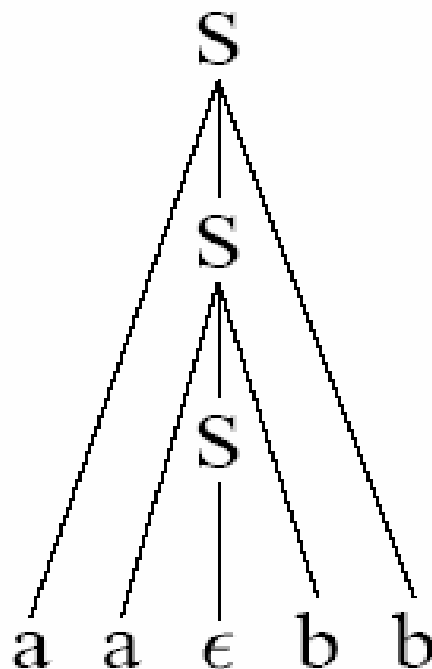
Фактически, безконтекстната граматика, моделираща $a^n b^n$ има само две правила. А те са:

$$S \rightarrow a S b$$

$$S \rightarrow \varepsilon$$

1.2. РАЗШИРЯВАЩАТА ЛЕМА

На фигура 4 е представено парсиращо дърво, което използва тази граматика, за пораждането на думата aabb:



Фигура 4. Безконтекстно дърво, пораждащо думата aabb



1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?

“Как бизнесът върви?” или “Как върви бизнесът?”

“Бягам на улицата.” или “Бягам върху улицата.”

“Слизам надолу.” или просто “Слизам.”

“между впрочем” или “между другото”

Разширяващата лема ни дава едно теоретично средство за разбирането на добре познатите аргументи, че Английският (или по-точно “набор от низове Английски думи, разглеждани като формален език”) не е автоматен език.



1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?

Първият такъв аргумент е даден от Чомски (Chomsky) /1957 г./. Той първи разглежда езикът xx^R , $x \in a, b^*$. x^R означава “обратният на x ”, така че всеки израз от този език е съставен от a -та и b -та низове, следвани от обърнат или “огледален образ” на низа. Този език не е автоматен.

След това Чомски доказва, че специално подмножество на Английската граматика е изоморфно с огледалния образ на езика. Дава да разгледаме следната синтактична структура, където S_1, S_2, \dots, S_n са декларативни изрази на Английски:

- ☐ Ако S_1 , тогава S_2
- ☐ Единият от S_3 или S_4
- ☐ Мъжът, който каза S_5 че пристига днес



1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?

Автоматните езици са затворени над операцията субституция или хомоморфизъм; това означава, че можем да преименуваме всеки от символите в горния израз. Да представим следната субституция:

ако $\rightarrow a$


тогава $\rightarrow a$

единият от $\rightarrow b$

или $\rightarrow b$

с други думи $\rightarrow \varepsilon$

Ако сега приложим това заместване за горния израз, ще получим сления резултат: $abba$



1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?

Този израз има тъкмо огледалната стойност, която показме по-горе и не се прихващаше от крайни автомати. Ако предположим, че ако (if), тогава (then), едiniqt от (either), или (or) могат да се вграждат произволно, то тогава Английският е изоморфен на xx^R , $x \in a, b^*$, и следователно не е автоматен език.

Партий (1990) дава второ доказателство, че Английският не е автоматен език. Това доказателство е основано на известно множество от изречения с централно-вграждани структури (Ингве, 1960).



1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?

Ето примерен вариант на тези изречения:

Котката обича риба-тон.

Котката, която кучето преследваше, обича риба-тон.

Котката, която кучето, която плъхът ухапа обича риба-тон.

Както бе с или..или/или изреченията по-горе, така и тези изречения стават по-трудни за разбиране, придобивайки по-сложна структура.

След като всяко пряка обработка на езика трябва да има свързан към нея глагол, то тези изречения са от вида:

(пълнен член съществителното)ⁿ (преходен глагол)ⁿ⁻¹ харесва риба-тон



1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?

Идеята на доказателството ще да покаже, че изречения от тези структури могат да бъдат произведени чрез сечение на английски с регулярен израз. След това ще използваме разширяващата лема да докажем, че резултатният език не е автоматен.

За да построим елементарен регулярен израз, който можем да пресечем с английския език за да построим тези изречения, дефинираме регулярни изрази за групата на съществителните (А) и глаголи (Б):

$A = \{\text{котката, кучето, плъхът, слонът}\}$

$B = \{\text{преследва, ухапа, възхити се, изяде, помагам, ...}\}$



1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?

Сега ако вземем регулярния израз A^*B^* харесва риба тон и направим сечение с английския (считайки го, като множество от низове), резултатният език би бил:

$$L = x^n y^{n-1} \text{ харесва риба тон, } x \in A, y \in B$$

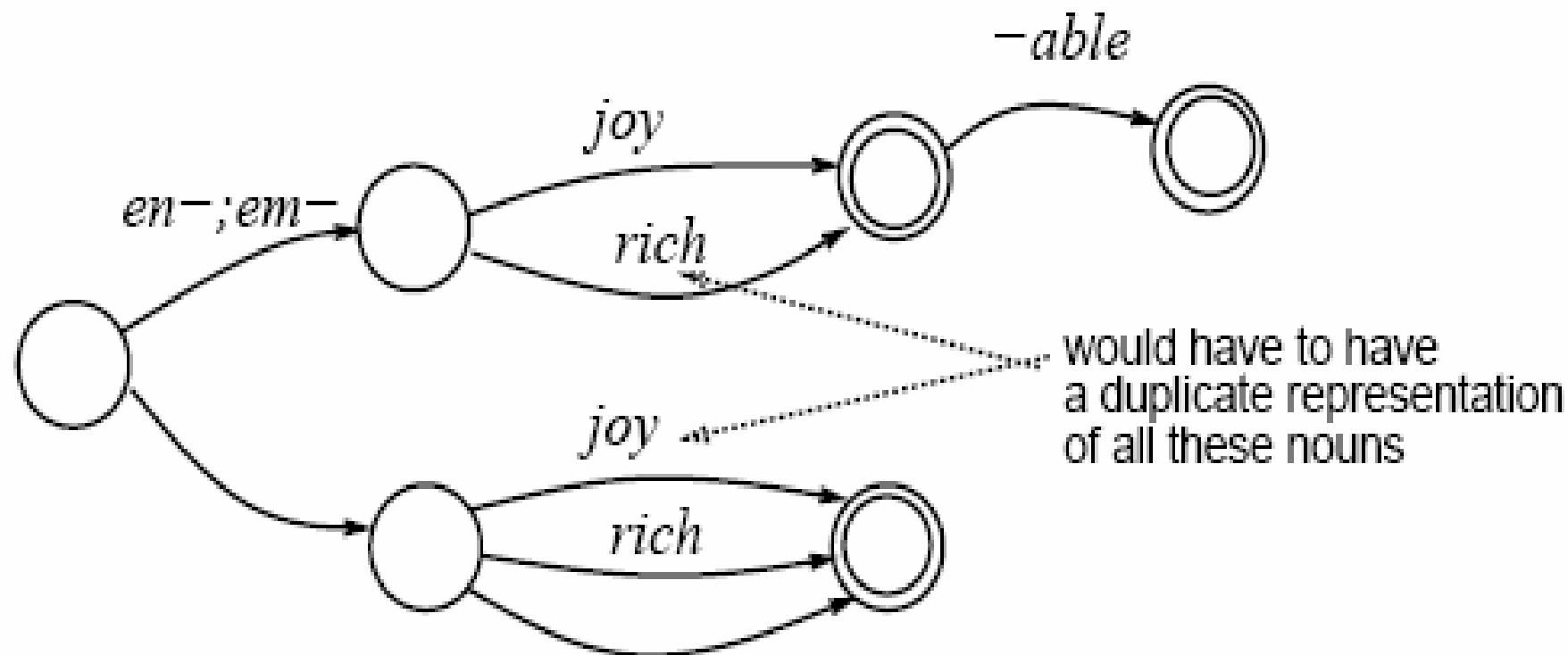
Резултатният език L може да бъде представен като не-неавтоматен, чрез лемата. След като сечението на английския с автоматен език не е автоматен език, то тогава английският също не може да бъде автоматен език.



1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?

Двата аргументи, които разгледахме досега са базирани на английския език. Съществуват и аргументи против крайната природа на английския, основани на английската морфология. Тези морфологични аргументи са различен тип аргументи, защото те не доказват че английската морфология не може да бъде автоматна, само че безконтекстния модел на английската морфология е много по-елегантен и прихваща някои полезни описателни генерализации.

1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?



Фигура 5. Част от машина с крайни състояния за английската морфология (Пост-Проут, 93г.)



1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?

Горният морфологичен факт е лесен за изразяване в безконтекстна граматика: нека предположим, че –en се прилага към конкретен клас от прилагателни и съществителни. Тогава трите безконтекстни правила са достатъчни за да моделират този феномен:

Глагол -> en- Прилагателно

Глагол -> en- Съществително

Прилагателно -> Глагол –able



1.2. АВТОМАТНИ ЛИ СА АНГЛИЙСКИЯТ И ДРУГИТЕ ЕСТЕСТВЕНИ ЕЗИЦИ?

Няколко учени оспорват, че английското съгласуване на числа не може да се прихване от автоматна (даже и без-контекстна) граматика.

Ето и тяхната приложена автоматна (с десен контекст) граматика, която моделира тези изречения:

$S \rightarrow \text{За кой проблем каза учителят че } T$

$S \rightarrow \text{За кои проблеми ти каза учителят че } U$

$T \rightarrow \text{тя помисли } T \mid \text{ти помисли} \mid \text{беше нерешим}$

$U \rightarrow \text{тя помисли } U \mid \text{ти помисли } U \mid$



1.3. БЕЗКОНТЕКСТЕН ЛИ Е ЕСТЕСТВЕНИ- ЯТ ЕЗИК?

В предишния раздел оспорвахме, че английският (разглеждаме като множество от низове) не изглежда като автоматен език. Следващият очевиден въпрос е да се попита дали английският е безконтекстен език. Този въпрос има интересна история: редица от опити за доказване на английската безконтекстивеност са били публикувани; и всички с изключение на два са били опровергани след публикацията си.



1.3. БЕЗКОНТЕКСТЕН ЛИ Е ЕСТЕСТВЕНИ- ЯТ ЕЗИК?

Един от тези 2 верни (или поне все още неопровергани) аргумента произлиза от диалект на Швейцаро-Германски.; другият от морфологията на Bambara, северозападен език говорещ се в Мали и съседните и страни.

И двата аргумента и повечето от опроверганите, използват факта, че следните езици и тези които имат близки свойства не са безконтекстни:

$$\{ xx \mid x \in \{ a, b \}^* \} \quad (12.1)$$

Този език се състои от изречения съдържащи два идентични слети низа. Следващият сроден език също не е безконтекстен:

$$a^n b^m c^n d^m \quad (12.2)$$

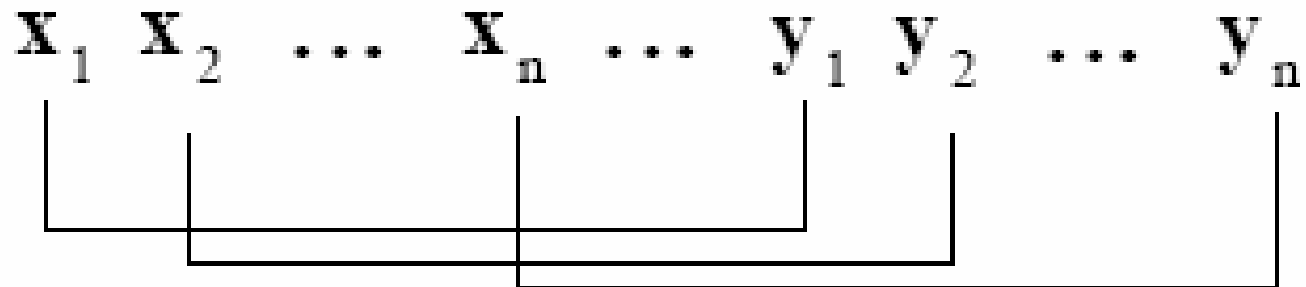


1.3. БЕЗКОНТЕКСТЕН ЛИ Е ЕСТЕСТВЕНИ- ЯТ ЕЗИК?

Опитите да се докаже че естествените езици не са подмножество на безконтекстните езици се осъществяват, като се покаже, че естествените езици имат свойствата на тези xx езици наречени пресичащи се зависимости. При пресичащите се зависимости, думи или по-големи структури се свързват от ляво на дясно, както е показано на фиг.6. Един език, който като произволно дълга пресичаща се зависимост може да бъде наложен към xx езици.

1.3. БЕЗКОНТЕКСТЕН ЛИ Е ЕСТЕСТВЕНИ- ЯТ ЕЗИК?

Фигура 6. Схема на пресичаща се зависимост





1.3. БЕЗКОНТЕКСТЕН ЛИ Е ЕСТЕСТВЕНИ- ЯТ ЕЗИК?

Сполучливото доказателство, самостоятелно предложено от Шибер (1985), показва че диалект от Швейцаро-Германски говорим в Цюрих има пресичащи се ограничения, които правят някои части от този език еквивалентен на не безконтекстния език $a^n b^m c^n d^m$.

Първо той отбелязва, че Швейцаро-немския диалект, позволява глаголи и техните аргументи да бъдат подредени в пресичайки се.



1.3. БЕЗКОНТЕКСТЕН ЛИ Е ЕСТЕСТВЕНИ- ЯТ ЕЗИК?

Ако предположим, че всички примерни клаузи посочени по-долу са предшестват от низ “Жан казва това”:

(12.3) ...помогнахме на Ханс да боядиса къщата”

Забележете пресичащата природа на семантичната зависимост. И двете съществителни предхождат глаголи, като Hans е аргументът на help (помагам), докато The House (къщата) е аргумент на paint (боядисвам). По-нататък има cross-serial dependency м/у глаголите и съществителните. Help изисква дателен падеж, Hans е в дателен падеж, докато paint приема винителната форма и the house е във винителен падеж.



1.3. БЕЗКОНТЕКСТЕН ЛИ Е ЕСТЕСТВЕНИ- ЯТ ЕЗИК?

Нека имаме един регулярен израз R . След като е регулярен израз (в него може да видите евентуално само конкатенация и звездата на Клийни), трябва да се дефинира и автоматен език и така ще можем да пресечем R с Швейцаро-немския диалект и ако резултатът е безконтекстен, то и диалекта ще е безконтекстен.

Оказва се че Швейцаро-немския диалект изисква броя на глаголите изискващи обекти в дателен падеж трябва да е равен на броя на съществителните фрази в дателен падеж. Подобна е ситуацията и за винителния падеж. И произволен брой от глаголи може да се появят в подчинена клауза от този тип (за изпълнение на ограничения).



1.3. БЕЗКОНТЕКСТЕН ЛИ Е ЕСТЕСТВЕНИ- ЯТ ЕЗИК?

Това означава, че резултатът от пресичането на този автоматен език с диалекта е език със следната форма:

$$wa^nb^mxc^nd^my$$

// Който не е безконтекстен!

И така можем да заключим, че Швейцарския немски диалект не е безконтекстен.



1.4. СЛОЖНОСТТА НА ИЗЧИСЛЕНИЯТА

Тази секция, в която ви запознаваме с обработката на естествените езици и обобщаваме аргументите на Barton (1987), може и никога да не види бял свят:

1. поддържаните лексикални и съгласувани особености са неясни над усиления инфинитив – дългите изречения причиняват проблеми при разпознаването на текстове на естествени езици.

2. второ ниво – прави се морфологичен разбор (или еднакво определен между лексикалната и повърхностна форма) е също така пълна обработка на естествения език.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Проницателният читател може да е забелязал, че много от изреченията които бяха използвани за оспорване на не автоматната природа на английския език (като централно-вложени структури са всъщност доста трудни за разбиране). Ако говорите Швейцаро-немски диалект (или ако имате приятел който го говори) ще забележете, че дългите взаимосвързани изречения са също толкова трудни за разбиране.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

В действителност както Пулум и Газдар (1982) отбелязват:

“... точно тези типове конструкции, които се отличават в различните доказателства че английският не е безконтекстен, изглежда масово създават затруднения в човешката система за обработка...”

Това ни навежда на второто значение на понятието: многозначност. В предишната секция говорихме за многозначността на езика. Тук се обръщаме към въпрос, който е колкото психологически, толкова и изчислителен: сложността на езика на човешката изразяване.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Много неща могат да направят едно изречение трудно за разбиране: сложно значение, крайно двусмислени изречения, използването на редки думи и лошият почерк са само някои от причините. Но има един особен тип сложност (често наричана синтактична сложност), която носи и интересна релация към сложността на формалния език от предишната секция. Това са изречения, чиято заплетеност се получава не е от редки думи или трудни значения, а от комбинацията от синтактична структура и ограничения в човешката обработка.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

(Тук ще използваме # за обозначение на изречения, които предизвикват най-високо затруднение при обработка). При всеки случай (ii), примерът е значително по-сложен отколкото (i) примерът:

1. (a) Котката обича риба тон.

(b) # Котката, която кучето, която плъхът ухапа обича риба-тон.

2. (a) Ако когато плаче бебето, майката се разстройва, бащата би помогнал, за да може да почива бабата.

(b) # Понеже ако плаче бебето, майката се разстройва, а бащата би помогнал, за да може бабата да почива спокойно.

3. (a) Детето повреди снимките, които бяха заснети от фотографа, който професорът срещна на партито.

(b) Снимките, които фотографът, който професорът срещна на партито, засне бяха повредени от детето.

4. (a) Фактът, че работникът, който управителят нае, краде офис оборудване, тревожеше администрацията.

(b) # Администрацията, която фактът, че работникът краде офис-оборудване, тревожеше управителя.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Най-ранните работи в/у изречения от този тип забелязват, че те всички включват вмъквания или централни-вграждания (Чомски 1957, Юнгве 1960, Чомски и Милер 1963, Милер и Чомски 1963). Това означава, че те всички съдържат примери където синтактична категория А е съв-местена в рамките на друга категория В, и заобиколени от други думи (Х и Y).

[В Х [А] Y]



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Най-ранните работи в/у изречения от този тип забелязват, че те всички включват вмъквания или централни-вграждания (Чомски 1957, Юнгве 1960, Чомски и Милер 1963, Милер и Чомски 1963). Това означава, че те всички съдържат примери където синтактична категория А е съв-местена в рамките на друга категория В, и заобиколени от други думи (X и Y).

[B X [A] Y]

Във всеки от примерите по-горе, част (i) има 0 или едно вграждане, докато част (ii) има две и повече вграждания.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Можем ли да обясним трудността на тези вложени структури, като просто кажем, че те са граматически неправилни в Английският език? Отговорът изглежда е НЕ.

Структурите, които са използвани в сложните изречения в 1-3(ii) са същите тези използвани и в по-лесните изречения в 1-3(i).

Разликата м/у елементарните и комплексни структури изглежда се основава в/у броя на вгражданията. Но няма естествен начин да се напише граматика, която позволява N-вграждания, но не и N __ вграждания.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Наистина изглежда че случаят на множествено вградени структури от този род са също толкова трудни и в други езици. Например японският позволява клауза за единично вграждане, но добавянето на допълнителна клауза прави изречението необработваемо.

(Купър 1976, Бабионишев и Гибсон 1996):

(12.5) Ani-ga	imooto-o	ijimeta.
по-голям брат-NOM	по-малка сестра-ACC	тормози
‘По-големият ми брат тормози по-малката ми сестра’		

(12.6) Bebiisitaawa	[[ani-ga	imooto-o
babysitter-TOP	[[older-brother-NOM	younger-sister-ACC
ijimeta]	to]	itta.
bullied]	that]	said
‘Детегледачката каза че по-големият ми брат тормози по-малката ми сестра’		



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

(12.7) #Obasan-wa	[[Bebiisitaa-ga	[[ani-ga
aunt-TOP	[[babysitter-NOM	[[older-brother-NOM
imooto-o	ijimeta] to]	itta] to] omotteiru.
younger-sister-ACC	bullied] that]	said] that] thinks

‘Леля ми смята че детегледачката каза че моят по-голям брат тормози моята по-малка сестра’

Има няколко опита да се обясни комплектността на тези резултати, повечето от които са базирани на памет-та.

Например Ингве (1960) направи предположение че човешкият механизъм за обработка (парсер) се основава на ограничен по размер стек. Стеково-базиран парсер поставя непълни правила на фразеологични-структури в стека; ако многобройни непълни фрази са вградени, стекът ще съдържа регистър за всяко от непълните правила.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Разширение към този модел (Милер и Чомски, 1963) предлага че по-специално самовложените структури са от особена трудност. Една самовложена структура съдържа синтактична категория A вградена в друг пример на A и заобиколена от други думи (X и Y).

$[A\ X\ [A]\ Y]$

Такива структури може да са трудни, защото парсер, базиран на стекове може да обърка две копия на едно правило в стека.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Проблемът със самовграждането е също така естествено моделиран с activation-based model, които могат да имат само едно копие на конкретно правило.

Един от проблемите е, че има много синтактични сложни ефекти, които не могат да се обяснят чрез тези модели. Например има значителна разлика в сложността между изречения, които имат същия брой вграждания, такива като добре познатата разлика между субективни относителни клаузи и обектните относителни клаузи:

- (12.8) (a) [S Репортерът [S_ който [S атакува сенаторът]] призна грешката].
(b) [S Репортерът [S_ който [S атакува сенаторът]] призна грешката].



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Обектно-относителните клаузи са по-трудни за обработка, (измерено например по количеството време, което от-немат, за да ги прочетем (Форд, 1983) и други фактори; вижте Ванер и Маратсоре (1978) и Кинг и Джъст (1991) и Гибсън (1998) за обобщение. Различни изследователи са изказвали предположения за броя на различните фактори, които биха могли да обяснят тази разлика в сложността. Например Гибсън (1998) посочва, че обектните клаузи имат две съществителни, които се появяват пред всеки глагол.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Счита се, че обичайните канонични структури са по-лесни за обработка, отколкото обектните. Друг потенциален източник на трудност в тях, е че първото съществително (репортерът) има две различни тематични роли – агент в едната клауза и пациент в другата.

Друг факт срещу обикновения парсер, базиран на нотация върху синтактичната сложност идва от изречения, които имат двойно вградени релативни клаузи, но са все още обработваеми.



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Съгласуваността на неинтегриран обработка на езика е много по-сложна. За илюстрация в следващите примери на двойно вградени относителни клаузи и структури от литературата, двойно вградени относителни клаузи са обработваеми, защото най-вътрешната обработка на езика (И) не представя новооткрит обект.

(a) Книга, [която някой италианец, [не съм чувала за него] пиша] ще бъде публикувана скоро от MIT Press.

(b) Картината, [която фотографът, [който срещнах на партито,] взе,] прекара много добре. (Бевер, лична коресп. с Гибсон).



1.5. СЛОЖНОСТ ПРИ ОБРАБОТВАНЕТО НА ЕЗИКА ОТ ЧОВЕКА

Ранните предложения, че сложността на обработката на човешката реч е свързана с паметта, изглежда коректно на някои нива: в естествените и във формалните езици, се причинява от нуждата да запазим много неинтегрирани неща в паметта. Това е дълбокият впечатляващ резултат от обработката на езика. Но връзката между формалната и естествената сложност не е проста, както Ингве и други смятат.

Точно кои фактори играят някаква роля за сложността е тема на вълнуващи изследвания, които започват тепърва да се развиват.