

РЕЗЮМЕ

На монографията

LANGUAGE, CONCEPT FORMATION AND CHILD LANGUAGE ACQUISITION

An information modeling approach

(ЕЗИК, ФОРМИРАНЕ НА ПОНЯТИЯ

И

УСВОЯВАНЕ НА ЕЗИКА ОТ ДЕЦАТА

Подход на информационно моделиране)

от

Велина Славова

ISBN-978-619-245-221-6

© Автор: Велина Тодорова Славова, 2022 г.

Издателство "Професор Марин Дринов"

на Българската академия на науките

Вместо увод

Никой не знае точно колко са езиците на света. Един авторитетен източник, който е събрал данни от цял свят, The Ethnologue, посочва, че общият брой на езиците е 7 139 (през 2021 г.). Когато целта е да се разгледа езикът като способност, присъща на всички хора, трябва да се разгледа по общ начин пълният набор от езици. До момента науката не е намерила обяснение за това как се е появил езикът, защо има толкова много различни езици и каква е общата им основа, ако има такава.

В настоящата монография езикът се разглежда като система за обмен на информация. Представеният тук подход свързва в обща рамка разсъждения от научните области на езика, когнитивната наука и информатиката. Подходът в тази монография разглежда езика от нивото на високоорганизиран феномен до нивото на съществени детайли. Стъпките на моделиране започват от общ модел на езика като когнитивен феномен, за да достигнат до модел на обработка на информация при формирането на понятия. Целта е да се проследят необяснени досега взаимовръзки между езика и менталните информационни единици и да се потърсят общи принципи, валидни на всички нива на обработка на субстанцията „смисъл“.

Работата е структурирана в четири части, представени накратко по-долу.

Част 1.

С помощта на Обобщена мрежа е съставен модел на разбиране на езикови съобщения, представен като обработка на паралелни информационни потоци. Подходът отразява и проследява информационните потоци, възникващи в когнитивния процес на извличане на смисъла на дадено съобщение. Моделирани са информационните потоци, известни в когнитивна наука като "отгоре-надолу" и "отдолу-нагоре". Тяхното взаимодействие имитира процеса на натрупване на нова смислова информация при използване на обратна връзка от вече наличната в паметта информация. Наличната информация и свързаните с нея обработки са моделирани като протичащи в две отделни "пространства на знанието" – знание на езика и знание за света. Анализът на потоците показва, че когнитивните процеси от висок порядък трябва да използват едновременно както лингвистичните, така и семантичните знания. Това поражда хипотезата, че обработката се извършва в "слято" пространство от лингвистични и семантични знания. Проследяването на входовете и изходите в мрежата води до заключението, че за да се изгради мисловен образ на едно изречение е необходима операция "Семантично сливане", водеща до получаване на обобщени семантични образи в съответствие както с граматичните правила, така и със семантичните връзки между участващите понятия. Трасирането на специфични езикови примери с помощта на Обобщената мрежа потвърждава предположението, че образите на семантичните единици, участващи в съобщението, се подлагат на Семантично сливане и че това определя мисловното решение относно смисъла на изказаното.

Предложената операция „семантично сливане“ предизвиква нататък изследване на правилата за структуриране на изречението на синтактично ниво с помощта на

лингвистичен модел, основан на операцията Merge, дефинирана в синтаксиса. Анализирането на известния модел на рекурсивна граматика на Чомски довежда до получаването на покриващо дърво на Фибоначи, свързващо само елементите на базовата синтактична структура – Субект, Обект и Реципиент, известни в лингвистиката като "аргументи" на изречението. Полученото „аргументно дърво на Фибоначи“ отразява специфичните операции Merge и Type-shift, както те са дефинирани в лингвистиката. Демонстрирано е, че предложеното формално представяне обяснява един необяснен досега факт от лингвистиката, а именно, че броят на аргументите на фразата във всички езици е най-много три. За разлика от съществуващия модел на синтактично дърво, дървото на Фибоначи удовлетворява както факти, произтичащи от лингвистиката, така и принципа за ограничен когнитивен ресурс, познат в когнитивната наука.

Показано е, че последователното сливане отдолу нагоре на семантичните Обекти (Entities) по това дърво, разглеждано като оператор за получаване на по-голяма смислова единица, "произвежда" еднозначни семантични образи на Субекта в изречението във връзка с другите аргументи, като по този начин изгражда ясен семантичен образ на фразата. Предлаганото дърво е оптимално както по отношение на броя на стъпките на сливане, така и по отношение на броя на Обектите, участващи в изграждането на смисъла на изказаното. Направено е предположение, че това дърво отразява принцип на обработка на информационните единици на предлингвистично ниво, т.е. – отразява принцип на "езика на мисълта". За да се провери консистентността на модела са проведени два експеримента, основани на семантично двусмислени граматични правила в българския език. Резултатите са в подкрепа на хипотезата, че менталната интерпретация на базовия синтаксис на фразата се осъществява като спонтанно "изчисление" в съответствие с аргументното дърво на Фибоначи.

Според лингвистичния модел, предложен от Soschen, семантичните Обекти, участващи в "изчислението" на фразата, преминават през операция Merge за да бъдат включени в мисловната конструкция на фразата. Това поражда хипотезата, че формирането на понятията се подчинява на подобен процес със сходни изисквания за оптималност. Тези разсъждения водят до по-нататъшно подробно изследване на връзката между езика и мисълта при разглеждане на научни резултати в множество области.

Част 2.

Във втората част се предлага когнитивен модел на езиковата способност. Въз основа на широк анализ на съвременни източници от когнитивна наука, науки за мозъка, мозъчни образни техники, еволюция на езика, лингвистика, усвояване на езика и др. са представени основните пътища на разсъждение при изграждането на модела. Избрано е невробиологично съответствие на термина "концепт" като измерима кортикална активация. Концептите се разглеждат в контекста на известната в психологията теория за "grounded cognition", т.е. като произтичащи от възприятията. Като основна рамка се използва моделът, предложен от Barsalou и широко приет в когнитивната психология,

според който осмислянето на света се извършва от гледната точка на действащо лице в околната среда.

Анализът на резултатите от специализираните източници в областта на мозъчната активност с fMRI води до заключението, че информационните единици се представят и организират като ментални образи, отразяващи гледната точка на самата действаща биологична система, т.е. от първо лице и не съответстват напълно на представеното в световноизвестни онтологии, например WordNet. Разработен е, на база на това наблюдение, когнитивен "Аз-центричен" модел на езиковата способност. Моделът разглежда източниците на информация на четири нива – от възприятието и емоциите през концептуализация, към структурирано мислене и комуникация на мислите. Въз основа на направения анализ на различни съвременни източници, в този модел се приема, че информацията, необходима за изграждането на концепта за Аз, е генетично и биологично заложена и че този първичен концепт определя както начина, по който информацията за света се представя вътрешно, така и изграждането на ментални образи на случващото се, в които Азът е естествен агент. Този когнитивен модел е обобщение на предложеното аргументно дърво на Фибоначи и на "Аз-центричния" начин на създаване на информационни единици. Моделът е подкрепен с анализ на синтактичните структури на изречения от голям корпус с детска реч. Анализът показва, че детските изречения се развиват в съответствие с нарастващи по сложност конструкции на аргументното дърво на Фибоначи, които постепенно се появяват в детските фрази от 11-месечна до четиригодишна възраст. Първата употреба на всяка следваща по сложност аргументна конструкция винаги е със самото говорещо дете в ролята на подлог в изречената фраза.

Част 3.

В третата част е приложен емпиричен подход за изследване на процеса на формиране на концепти въз основа на статистически анализ на детската реч в началния етап на речевата продукция. Подходът е насочен към изучаване на биологичните принципи на вътрешно представяне на заобикалящия свят, тъй като децата не знаят език при раждането си и вродените механизми за класифициране и организиране на информацията подпомагат процеса на формиране на понятия при взаимодействието на детето със средата и езика. Представени са източниците на данни, както и процедурата, довела до събирането на много голямо корпус с данни от свободни диалози на деца на възраст от 9 месеца до 5 години, проговарящи английски или френски език. Статистическият анализ започва с откриването на разлики между половете в употребата на категории съществителни имена, като категориите са изведени чрез прилагането на когнитивния Аз-центричен модел, предложен в предходната част. Откриват се статистически значими разлики в употребата на пет от 14-те предложени категории, като разликите са в унисон с общоприети теории за спецификата на активностите на мъжете и на жените в древните човешки общества. Този резултат подсказва, че разсъжденията, следвани в когнитивния модел, водят до смислен резултат.

Направен е статистическия анализ на процеса на усвояване на двата езика, който показва, че когато думите на децата се класифицират според съществуващите в двата езика системи за части на речта, процесите на усвояване на английския и френския език не показват почти никакво сходство. Приложеният факторен анализ, обаче, показва, че, напротив, двата езика се усвояват по идентичен начин, както става ясно от първата базова компонента на статистическото пространство на употребата на части от речта във времето на усвояване. Предположено е, че тази статистически висока степен на сходство (0,905) се дължи на универсалната когнитивна основа на процеса на усвояване на езика и следва да се обясни на някакво по-дълбоко ниво, свързано с формирането на концепти и с "изчислителната" сложност на концептите като информационни единици. Това изисква концептите да бъдат представени в класове, които отразяват както значението, така и когнитивната "изчислителна" сложност.

Въз основа на когнитивния Аз-центричен модел се предлага начин, по който информацията е вътрешно представена в класове от концепти. Предложеният набор от класове се основава на разсъждения за различните по тип информация, необходими за управлението на автономна система, която може и трябва да действа в своята среда. Получените по този начин класове почти напълно съвпадат с известна теория за семантично подпомагане, предложена от Пинкър в когнитивната лингвистика. Нататък, статистическият анализ показва, че когато частите на речта в двата езика се разглеждат като принадлежащи към предложените класове концепти, процесите на усвояване на двата езика от децата са много сходни. За два класа обаче – Отношения и Обстоятелства – статистическият анализ показва наличието на различни пропорции на употреба от двете езикови групи деца. Предположено е, че смисълът, изразяван от деца на една и съща възраст е еднакво сложен и че разликите се дължат на особеностите на езиците. Това води до хипотезата, че има и други фактори, които обясняват почти пълното съответствие на първите базови компоненти на двете статистически пространства, които, явно, отразяват единната когнитивна основа за усвояването на двата езика. Както е известно, в изследвания период от детско развитие когнитивните способности нарастват и сложността на решаваните когнитивни задачи расте вследствие на увеличаването на когнитивните ресурси. Това налага да се анализира процесът на формиране на концепти от гледна точка на неговата когнитивна сложност, разглеждана в "изчислителен" контекст.

Част 4.

В четвъртата част е предложен модел за обработка на информация при формирането на концепти. Разсъжденията се основават на основни факти в неврологията. Важното предложение е, че информационните единици се представят като резултат от сравняването на сигнали от два източника – тези, идващи от биологичната система, включително паметта, и тези, идващи от околната среда. Представен по този начин, процесът на създаване на информационни единици се представя с двоично дърво, възлите на което моделират операцията „информационно сливане“, а ребрата – пътища на предаване на нервните сигнали. Така представено, дървото отговаря на условията на теорема, доказана от Хорибе, относно дърветата на Фибоначи и тяхната оптималност.

Тази стъпка на моделиране води до представяне на процеса на създаване на информационни единици чрез дърво на Фибоначи, математически доказано като структура с минимална цена (изразена в термините на тегла на всички пътища от листите до корена) при максимална ентропия, изразена на листата. Приложено в контекста на модела, това дърво описва минимални разходи на биологични и енергийни ресурси за получаване на максимално количество класифицирана информация, когато е "получило" като сигнал максимална входна ентропия.

Предложените класове на концепти са съотнесени към нивата на това дърво в зависимост от тяхната степен на абстрактност. Полученото представяне на дървото на Фибоначи предоставя метод за оценка на сложността на класовете концепти по начин, аналогичен на изчислителна сложност. Предложена е мярка за "когнитивна" сложност на всеки клас, която изразява усилията за формиране на концепт като брой операции по свързване на сигналите, произтичащи от околната среда, с тези, произтичащи от биологичната система.

Сложността на предлаганите класове е заместена в адитивен израз, представящ общата когнитивна сложност на това, което е изказано в диалозите на децата. Когато стойностите за честотите на използване на класовете, произтичащи от данните в корпуса се заместят, този израз за двата езика се развива идентично с във времето на усвояване. Корелацията е 0.933. По метод Монте Карло е отхвърлена хипотезата, че тази висока корелация е получена по случайност. Общата "сложност на концептите" използвани в детските изрази и на двата езика е резултат от прилагането на модела на сложност на концептите. Това показва, че наблюдаваният процес на усвояване на езика се определя от нарастването на когнитивните ресурси с възрастта на децата и че както англоговорящите, така и френскоговорящите деца усвояват и използват концептите в зависимост от тяхната когнитивна сложност, при ограничение от еднакво нарастващия за двете групи когнитивен ресурс. Различията в двата езика в интензивността на употребата на два от класовете се дължат, явно, на различия в езиковата практика и се появяват чак след тригодишна възраст, при това – при използвана еднаква обща когнитивна сложност при двете групи.

Статистическият показател за сложността на равнище на изказаните изречения също се развива във времето по идентичен начин в двата езика. Това от своя страна означава, че сложността на предаваните "идеи" е еднаква и за двете групи деца и че тя се определя от наличните когнитивни ресурси. Усвояването на английски и френски език, както на ниво сложност на концептите, така и на ниво сложност на изказаните идеи, се развива еднакво в рамките на наличните и нарастващи когнитивни ресурси. Разликите възникват поради различните езикови средства, използвани за изразяване на едно и също нещо – мисълта, която трябва да бъде предадена.

Като се има предвид, че получените модели, отразяващи както нивото на синтактичната структура, така и нивото на формиране на концепти, се описват формално с една и съща структура – дърво на Фибоначи, може да се предположи, че създаването на информационни единици и формирането на по-големи информационни

единици се подчиняват на сходни принципи, които отразяват оптимално използване на ресурсите.

Общ научен принос – самооценка

Работата обединява ключови теории и резултати от областите на лингвистиката, когнитивната наука и психологията, за да изгради поетапно информационен модел на начина, по който се придобива и организира информация в биологичните системи. По аналогия на изчислителната сложност, предложена е мярка за когнитивната сложност на класове от концепти с различно ниво на абстрактност. Моделът, основан на резултати в областта на математиката, свързва придобитата и класифицирана информация с принципите на оптимално използване на биологичните и енергийните ресурси. Емпиричното потвърждаване на модела е направено чрез статистически анализ на голям обем данни от детска реч. Прилагането на модела води до представяне на процеса на усвояване на английски и френски език като един и същи процес. Това емпирично потвърждение свързва езиковия изказ на децата, представен чрез сложност на ментална обработка, с нарастващите им когнитивни ресурси. Доколкото е известно на автора, досега няма подходи, които да съчетават в един модел възникването и класифицирането на информация, принципите на оптималност и усвояването на езика, за да обяснят езиковата способност като корелирана с, и ограничена от когнитивните ресурси.

На вниманието на научното жури са представени рецензии на монографията от:

Красимир Атанасов

Професор по Комуникационна и компютърна техника и професор по математика,
Институт по биофизика и биомедицинско инженерство,
Българска академия на науките,
София, България

Енчо Герганов

Професор по психология и психолингвистика, Департамент по психология и
когнитивни науки, Нов български университет,
София, България

Васил Василев

Професор по изкуствен интелект и интелигентни системи, Факултет по информатика и
цифрови медии, Лондонски университет "Метрополитън",
Лондон, Обединеното кралство

Виталий Величко

Професор по информационни технологии, Институт по кибернетика,
Национална академия на науките,
Киев, Украйна

Ирина Блинникова

Професор по когнитивна психология, Факултет по психология,
Московски държавен университет "Ломоносов",
Москва, Русия

Забележка – Рецензията от проф. Блинникова пристигна след предаване на монографията за печат.