

Владимир А. Карпов

ИЗОМОРФИЗМ ЗНАНИЙ О ЯЗЫКЕ И МИРЕ

Не секрет, что большая часть знаний о мире представлена средствами естественного языка и меньшая – формульно. Поэтому к первозадаче проблематики искусственного интеллекта я отношу формализацию знаний о языке и мире. Насколько возможна эта формализация можно судить по данному краткому сообщению.

Своеобразной революцией в химии был переход от брутто-формул к структурным формулам, что далее позволило нейрофизиологам получать качественно новые результаты в исследовании деятельности мозга: "Современная нейрофизиология на основе нейрохимических молекулярных процессов доказывает, что жизненная значимость отдельных событий представлена в мозгу даже в специфических химических процессах мозга, которые как бы усекают "шаги" этих жизненно важных событий. Так, например, мы имеем различную химию страдания, тоски, страха, радости и других существенных эмоциональных переживаний и событий в жизни животных и человека" [Анохин, 1978 : 33].

Такое положение в двух связанных областях не может не подталкивать интерес лингвистов к решению вопроса – каким образом может храниться в мозгу (в памяти) визуальная буквенная и слуховая информация. Если состояния имеют различную химию, то материальные единицы – слышимые, видимые и написанные слова представляют собой явно структурированные образования, главной особенностью которых являются реализации их конвергентных и дивергентных блоков – ср. Мама – Лама – Дама – Рама – Сама – Гама и т.п.

До выяснения химических эквивалентов звукового слова еще далеко, но данные нейрофизиологов показывают, что работать в этом направлении следует и следует искать универсальные способы.

Так, в работе "Психическая форма отражения действительности" П.К.Анохин пишет следующее: " Передача информации в живых и технических объектах обычно происходит с исключительно большим количеством специфических звеньев, однако она подчиняется одному важному закону: между начальным и конечным звеном этой передачи должна быть точная и адекватная информационная эквивалентность" [там же, 338]. Далее для подтверждения своей мысли о необходимости существования информационного эквивалента сигнала автор ссылается на опыты Уивера и Брея и пишет: "Если на слуховой нерв, проводящий потоки нервных импульсаций, в которых эквивалентно закодирован звук "а", наложить электроды и если получающиеся при этом биотоки нервных волокон преобразовать соответствующим образом, то на репродукторе можно услышать именно звук "а"" [там же, 339].

Материальные объекты различных предметных областей существуют в той или иной форме, предполагающей определенную внутреннюю структурированность. Вне зависимости от конкретной предметной области эти структуры при сопоставлении обнаруживают совпадение, сходство, подобие, в пределе тождество (как разновидности симметрии) и несовпадение, несходство, отсутствие подобия, нетождественность (как разновидности асимметрии). При частичном совпадении и одновременном частичном несовпадении признаков

сопоставляемых объектов принято говорить о диссимметрии. По мнению П.Кюри диссимметрия обнаруживается в любом явлении и причинах его создающих.

Таким образом проблема упаковки (формы, вида) того или иного объекта-системы в рамках предметной области выходит сегодня на новый симметрично-асимметричный теоретический уровень. Особенно велики достижения в этом направлении, полученные в химии, кристаллографии, биологии. О важности этих результатов можно судить по следующему фрагменту: "Дело в том, что так или иначе выявление видов симметрии конкретных биообъектов связывается с выявлением способов упаковки тех или иных компонентов в эти биообъекты... Понятно, что выявление видов биологических упаковок поможет глубже понять сущность жизни" [Урманцев, 1974: 222]. Если исходить из этого тезиса, то выявление способов упаковки языковой информации поможет глубже понять сущность языка.

Химики, придя к формульному языку представления знаний о своей предметной области, получили возможность приращивать знания лишь за счет их системного представления. Такая ситуация подтолкнула нас к отысканию возможностей формульных представлений сначала языковых изомеров, а затем и всей лексики, что привело к вполне оптимистическим нетривиальным результатам.

Суть феномена изомерии заключается в том, что сложные конструкторы могут быть одинаковы по составу элементов нижележащего уровня, но благодаря разному порядку следования этих элементов, конструкторы имеют различные свойства. Простейшие примеры лексической изомерии (существует и синтаксическая изомерия) – *араб-арба-раба-бара*; *звонок-звонко*; *модернизируете-дезориентируем* и т.п. позволяют увидеть не только лексико-грамматическую разницу и (или) сходство слов-изомеров, но и показывают связи феномена изомерии с феноменами другого рода, в частности, с омонимией, поддерживающей изомеризацию системы в целом.

Языковая изомерия, существование которой было предсказано Ю.А.Урманцевым [Урманцев, 1974], вскоре была доказана в работе [Карпов, 1992] и исследована в кандидатской диссертации С.Г.Барбук, где был вскрыт и описан ряд грамматических и семантических особенностей этого фундаментального явления, отмеченного у объектов неживой и живой природы [Барбук, 2000].

В цикле работ 2001 года впервые сделана попытка создания формульного языка для представления лексики любого языка мира в статике и динамике [Карпов 5, 6, 7, 8]. Одновременное существование признаков симметрии и асимметрии у языковых объектов, приводит к диссимметрии в языке-системе, представляющей по П.Кюри любое явление и обнаруживающаяся в причинах этих явлений. Конвергентно-дивергентные и дивергентно-конвергентные формулы и представляют диссимметрию. Доказано, что лексика русского языка может быть записана с помощью двух десятков стандартных формул конвергентного и дивергентного вида. Эти формулы представляют изомерную и неизомерную лексику, что позволяет говорить о теоретической полноте формульного языка-системы.

Аксиоматика формульного языка

Аксиома 1. Слова (лексемы или словоформы) можно рассматривать и представлять парами, что практически и делается в фонологии, грамматике, лексикологии.

Аксиома 2. В произвольно взятом естественном языке имеются словоформы одинаковой и разной длины. Более логично сопоставление слов одинаковой длины, т.к. при равенстве одного типа (симметрии) отчетливее выявляются неравенства (асимметрия); обоснование языковой симметрии см. [Карпов 9, 10, 11].

Аксиома 3. Произвольно взятая словоформа может быть структурно представима в терминах Н, С, К, где Н (начало), С (середина) и К (конец) словоформы.

Аксиома 4. Две произвольно взятые словоформы равной длины могут характеризоваться как конвергентные, дивергентные и конвергентно-дивергентные (дивергентно-конвергентные) в терминах Н, С, К и их комбинаторике, что в конечном счете приводит нас к представлению о симметричных (одинаковых, совпадающих) и асимметричных (неодинаковых, не совпадающих) началах, серединах и концах двух словоформ. При формальном анализе речь идет о буквах или фрагментах из букв, при содержательном – о морфемах.

Аксиома 5. Дивергенция (Д) может сменяться конвергенцией (К), а конвергенция сменяться дивергенцией, образуя ряды-цепочки двух видов:

- а. ДК, ДКД, ДКДК, ДКДКД, ДКДКДК, ДКДКДКД, ДКДКДКДК и т.п.;
- б. КД, КДК, КДКД, КДКДК, КДКДКД, КДКДКДК, КДКДКДКД и т.п.

Неспецифичность, простота и немногочисленность аксиом как исходных предпосылок для формульного языка, логичность построения формул как циклического развития, естественность - создают универсальность более высокого уровня – межсистемную, что следует понимать так, что лексика любого языка мира, представленная кириллицей или латиницей (для иероглифических языков), может быть упакована 2-3 десятками формул, к которым сводится все многообразие пар (миллионы словоформ).

Часть формул представлена в матрице. Черный кружочек формулы соответствует узловой, конвергентной, общей для двух слов точке (букве), или фрагменту из точек (букв), каждая из которых занимает одну и ту же позицию в слове. Белые кружочки представляют дивергентные точки или фрагменты.

Подчеркиваем, речь идет прежде всего о графических, графо-фонематических и фоно-графо-фонематических изомерах, хотя мы доказали существование семи и только семи типов изомерии, теоретически предполагаемых при исходных концептах в виде звука, буквы и фонемы (фонетические, графические и фонематические изомеры как базовые и четыре комбинаторных типа: фонографические, фонофонематические, графофонематические и фонографофонематические) [12]. Выяснение количественной представленности каждого из типов – предмет последующей отдельной работы, представляющий целое направление анализа языковой изомерии.

Возьмем пары словоформ-изомеров типа – **уход: худо**, где в первом слове рисунок ГСГС, а во втором обратные ему СГСГ (Г – гласная, С - согласная буква). Их запись друг под другом даже при одинаковом буквенном составе

уход опер удар анод икон ерик ажур
худо поре дура надо кино реки жару

после сравнения по позициям, занимаемым буквами, позволяет говорить о полном несовпадении позиций и присвоить таким парам номер формулы 0, представляющей предструктурные образования, когда все точки одного слова представляют оппозицию всем точкам второго слова полностью дивергентные словоформы объединяемые двумя параметрами состав и длина (для изомеров) или только длиной (для неизомеров)

Слова ОН и НО интересны тем, что начало первого совпадает с концом второго и конец первого совпадает с началом второго. Этому варианту со сдвигом мы присвоили номер 1. Общая конвергентная часть и различающаяся дивергентная в минимуме имеют единицу, максимум может достигать до 10.

Теперь рассмотрим слова АД и ОД. У них разные начала А и О и одинаковое завершение в виде Д. Несложно найти и обратную пару – ДА и ДО, где при одинаковом начальном символе Д не совпадают финальные точки А и О. Они будут представлять соответственно формулы 2 и 3. В формулах они не изомеры, а раздельно попадают в формулу 1 как изомеры – ср. АД и ДА, ОД и ДО.

Пара слов ТОК и КОТ имеет общий центр при разных началах и концах (формула 4), а в словах ДАМ и ДОМ, или ДАМ и ДЫМ, или ДОМ и ДЫМ мы находим одинаковые начала и концы при разных серединах (формула 5). Такие варианты совпадений и несовпадений обнаружены при длинах словоформ 2, 3, 4 буквы.

Рассмотрение динамики слов *кот* и *ток* (формула 4) в виде *котам:токам* показывает, что дивергенция статики в динамике превращается в конвергенцию, что превращает формулу 4 в формулу 6 и дает нам право далее строить формулы стандартным образом, добавляя конвергентные или дивергентные блоки. Поэтому матрица приводится в сокращенном варианте.

Последующие формулы наращивают длину единообразно (расхождением последней узловой точки или схождением разошедшихся точек). Формула внешне напоминает полимерную цепочку, где тиражируется многократно один и тот же фрагмент (в химии – линейное аннелирование). Так, формула 33 будет представлять восемь сцепленных формул 5, а 41-ая формула – 10 таких ромбиков.

Далее все четные формулы представляют дивергентные начала, а все нечетные – конвергентные начала слов. Матрица формул не случайно называется так, потому что дальнейшее порождение формул после третьей идет стандартным путем – через чередование конвергентных и дивергентных процессов, расширяющих формулы в длину.

Число зон дивергенции и конвергенции определяется для четных формул $(N : 2) + 1$, для нечетных $(N + 1) : 2$, оно же является и минимумом длин формул.

В матрице представлены наиболее нагруженные формулы с нулевой по 15). Можно полагать, что при анализе слов длиной более 12 букв могут обнаружиться заполнения 18 и 21 и более длинных формул (скорее всего неизомеров) – на белорусском материале найдены 24 и 26, но до автоматизации этого процесса дело еще не дошло, т.к. нет компьютерной версии всех словоформ русского языка. Проанализированы 50 тысяч изомерных пар и это дает основание для выдвижения гипотезы о том, что более сложные формулы могут вообще не реализоваться, так как конвергентные и дивергентные точки имеют тенденцию превращаться в блоки и вводить длинные слова в более простые по структуре формулы.

Матрица формул

0		1	
2		3	
4		5	
6		7	
8		9	
10		11	
12		13	
14		15	

Приводимые ниже примеры представляют: минимальный вариант заполнения формулы (необходимая для формулы минимальная длина в буквах) и расширенные варианты, включающие изомеры-словосложения. Они приводятся для того, чтобы показать, что усложнение морфологического состава слова хотя и ограничивает изомеризационные возможности системы, но не настолько, чтобы в ней полностью отсутствовали сложные слова изомеры. Пять самых наполняемых формул снабжены пометами в скобках. Отметим, что в отличие от предыдущих работ формула 0 и 1 имеют несколько иные рисунки, требуемые формализмом и перенумерованы: старая 0 равна новой 1 и старая 1 равна новой 0.

Формула 0: ишак – шика, уходил – ходули, ледовоз – довезло, шутоломе – отшумело, буквоед – обдувке, стекловар – кластеров и т.п.

Формула 1(пятое место по частоте): ял – ля, наморозят – тормозная, обусловлю – любослову, лаковаров – ворковала, курослеп – пролеску и т.п.

Формула 2: сои – оси, ламутка – мулатка, костоправ – простаков, краскомом – макрокосм, истмат – мастит, солевоз – лесовоз, разное – образное и др.

Формула 3 (четвертое место): жим – жми, исчезал – исчезла, медогон – медного, пимокат – покатым, паромер – промера, пароход – прохода и т.п.

Формула 4 (второе место по частоте): нос – сон, машин – нишам, меломана – наломаем, салотопки – полосатик, сопромат – пастором, кривошип – прошивке и др.

Формула 5: акру – арку, пропрядут – пропудрят, самокатный – самотканый, краболовом – карболовом, полкруга – прогулка, автолавка – автоклава и т.п.

Формула 6 (третье место по частоте): бузе – зубе, вредного – дверного, небосклоном – оскобленном, тиходом – отходим, водоток – отводок и др.

Формула 7 (первое место по частоте): даму – дума, скарбнице – скребница, ветреного – ветрогоне, полметра – прометал, пилораме – полимера и т.п.

Формула 8: икону – окуни, забелели – лебезила, лежебок – желобке, белорус – лесоруб, полувалы – уплывало, куроедов – руководе, мышелова – вымешало и др.

Формула 9: гадам – годам, манерой – мареной, полуротными – полуторными, нарпитом – напортим, кролиководы – крокодилова, старпом – стропам и т.п.

Формула 10: догола – голода, постулата – отступала, полутон – лопотун, стооком – откосом, волкобой – лобковой, салотопом – остолопам и др.

Формула 11: вареги – верига, продавали – подварила, педсовета – подсевае, домовод – доводом, маловеры – мыловаре, порнограф – профорган и т.п..

Формула 12: ходокам – комодах, пантомима – тампонами, водоход – доходов, рисовод – сидоров, водонос – доносов, волопас – посовал и др.

Формула 13: бабником – бобинкам, столами – смотали, пристроилась – простиралась, разморенный – разномерный и ряд других.

Формула 14: заржавел – разжевал, катализа – закатаила, ренегатский – тегеранский, водополье – половодье, водоразборный – рыбаобразной и др.

Формула 15: натравило – нитровали, пановали – понавила, резиночка – розанчике и пока единственный пример словосложений: полусмерть – пульсометр.

Формула 16: спиртовали:приставило.

Формула 17: наводнение – невиданное.

Формула 18: косоворотка – молокососка (неизомеры).

Формула 19: понакидали – паникадило.

Формула 20: недоработал – велозаводам (неизомеры).

Формула 22: недоработали – велозаводами, безоговорочной – сероводородный (неизомеры).

Формула 24: безоговорочными – сероводородному (неизомеры).

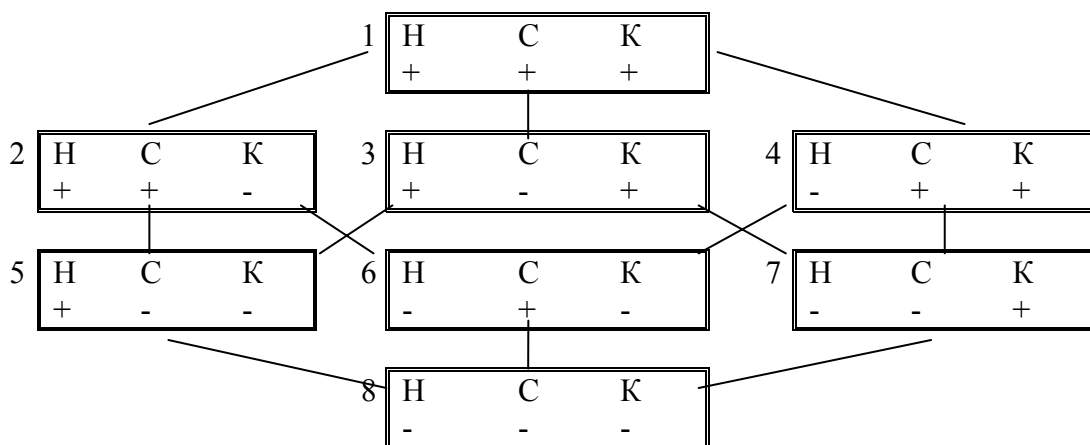
В структурном отношении зафиксировано увеличение длины слов, превышающее пиковые величины – от 9 до 18 букв (у сложных слов) при аналогичных пиках у текстовых слов 7-9 [13, 930].

Реализация большинства формул говорит о том, что словосложения подчиняются тем же самым законам создания наиболее плотных упаковок, описываемых некоторыми строками треугольника Паскаля. К ним мы приходим при анализе теоретически предсказуемого числа вариантов заполнения формул при последовательном увеличении длины слов от минимальной на 1 символ, когда растет число вариантов формулы, при этом вместо совпадающих/несовпадающих точек отмечаются фрагменты (2, 3, 4, 5 и более букв). Само же число вариантов формул определяется с помощью треугольных (1, 3, 6, 10, 15, 21,...), пирамидальных, или тетраэдрических (1, 4, 10, 20, 35,...) и других многоугольных чисел. А эти числа связаны с наиболее плотной упаковкой шаров. Анализ показал, что до длины 8 букв мы имеем дело с полным числом вариантов (плотные упаковки), с увеличением длины слов в формулах плотность снижается, доходя до единственного варианта. Так, для формулы 10 при длине 18 букв теоретически возможны 6188 структурных вариантов, рассчитанных по треугольнику Паскаля. Обнаружен пока только 1. Далее с увеличением длины изомерийные формулы вырождаются, т.е. становятся ненаполняемыми. Как было упомянуто выше – языковая материя становится настолько усложненной (обрастание изначально короткого слова префиксами, суффиксами - при словообразовании и присоединение окончаний при словоизменении), что исчезает возможность совпадения как таковая. Вскрытые особенности указывают и основные направления дальнейшей разработки

теории языковых упаковок: выделение плотного ядра и его анализ; выделение менее плотной периферии и выделение зоны неизомеризации.

Теперь обобщим эти формулы с точки зрения значимых концептов в виде: начало (Н), середина (С) и конец (К), т.к. начало слов в статике в динамике связывается с префиксацией, середина слов с увеличением числа корней при словосложении, а динамика конца слова с прирастанием суффиксов и окончаниями. Система в виде трехмерного куба дает теоретическую полноту в виде 8 подсистем. Плюс означает совпадение признаков, минус – несовпадение.

Рис.1. Структурное обобщение формул



Подсистема 8 должна содержать изомеры без совпадения начал, середин и концов. Эта ситуация соответствует формуле 0 (см. матрицу формул). Напоминаем, что совпадение следует понимать как нахождение одного и того же символа или фрагмента в одной и той же позиции равно как и несовпадение, ср. слова АСОМ и САМО, демонстрирующих несовпадение.

Подсистема 5 должна содержать изомеры с конвергентным началом. Такое условие соответствует формуле 3.

Подсистема 6 должна содержать изомеры с конвергентной серединой, что соответствует формулам 4, 8, 12, 16 и далее через 4 номера.

Подсистема 7 должна содержать изомеры с ковергентными концами, что соответствует формуле 2. Как видим, более простые формулы представляют более нежесткие некомбинаторные условия.

Подсистема 2 соответствует формулам 7, 11, 15, 19 и далее через 4 номера.

Подсистема 3 соответствует формулам 5 и 1 с оговоркой (начало одного изомера совпадает с концом другого изомера) для формулы 1.

Подсистема 4 соответствует формулам 6, 10, 14, 18 и далее через 4 номера.

Подсистема 1 требует совпадения начал, середин и концов изомеров в парах и соответствует формулам 9, 13, 17, 21 и далее через 4 номера.

Таким образом, все формулы поглощаются восемью теоретически заданными Н, С, К типами и естественно, если нас интересует одинаковая префиксация в изомерии, то мы будем исследовать лишь четыре подсистемы: 1, 2, 3 и 5. Изомеры без префиксов или с разными префиксами будут находиться в подсистемах 4, 6, 8 и 7. Кроме того, можно предсказывать - в какую из подсистем попадут все последующие по номерам формулы.

Далее следует рассмотреть количественно-качественное наполнение подсистем и исследовать статику и динамику явления, взаимосвязи формул и подсистем. Это необходимо сделать по той причине, что в одних случаях словообразование происходит в рамках одной формулы, ср. формула 3 – *весит-вести* и дериваты *свесит-свести, перевесит-перевести, отвесит-отвести* и т.п. В другом случае имеет место переход от одной формулы к другой, ср. *вес-сев* (формула 4), а *перевес-пересев, отвес-отсев, вывес-высев* и т.п. уже представляют формулу 7. При этом явны случаи поддержки изомеризации омонимией – **пересев** (от *пересесть* и от *пересеять*). Аналогичная картина отмечается и при словоизменении, когда изомеры – *перлит-претил* относятся к формуле 7, а *перлита-претила* к формуле 9; *локатор-рокотал* относятся к формуле 8, а *локатора-рокотала* - к формуле 10. При этом интерес представляют переходы внутри левых формул, внутри правых и переход левых формул в правые и наоборот (левые - четные формулы, правые – нечетные). У левых дивергентные начала и конвергентные или дивергентные концы, у правых же - конвергентные начала и дивергентные или конвергентные концы. Это служит показателем полноты в плане конвергенции и дивергенции.

Анализ более детальный касался одного из конкретных элементов сложных слов. Для этого был выбран наиболее омонимичный из них: "**ком**", представляющий усеченные слова – *командир, коммунистический, комиссар, командующий, комитет* и т.п. Он отмечен в сложных словах в правой позиции длиной от 4 (уком – уездный комитет) до 10 букв (замнаркоме – наркомзема). Оказалось, что из 31 зафиксированного в Обратном словаре (объем 121 тыс.) сложного слова с элементом –ком в правой позиции, изомеризованы 16 слов, чуть больше половины и этот элемент представлен в сложных словах с формулами - 0, 1, 2, 3, 4, 6, 7, 8, 9, 10 (приводим несколько наиболее интересных реализаций некоторых формул, напр., **горком** – громко – ф 3, **рудкома** – дураком – ф 4, **наркомам** – карманом, **партком** - тропкам – ф 6, **горком** – громок – ф 7, **райком** – маркой – ф 8, **партком** – порткам – ф 9). Сходную картину наблюдаем и тогда, когда элемент "**ком-**" находится в первой левой позиции: **комвуз** – звуком, **комбата** – табаком, **компартии** – притоками и другие. О широком охвате процессом изомеризации словосложений говорит и тот факт, что возможны изомеры только внутри сложных слов (*водоскатом – водостокам – скотоводам; иноверку - военруки, наркомзема – замнаркоме* и др.), а также изомеры словосложения и обычные слова. При этом одному сложному слову может соответствовать несколько изомеров – ср. **ВОДОЛЕЙ** – *деловой, долевого, ледовой, Володей* (имя собственное) и *володей* (от *володеть*). Другими словами, изменение структуры элементов сочетаний, сжимаемых в сложное слово, не отражается на формульных представлениях. Другое дело - число изомеров-словосложений. Их число на фоне слов, созданных аффиксацией, резко падает.

Упомянутую выше универсальность формульного языка мы проверили на белорусском (близкородственном – восточная подгруппа) - 20 формул, в болгарском (южно-славянском) - 13, в английском языке обнаружено 11 формул (языки аналитические) и в итальянском языке обнаружены 20 формул. Таким образом, можно ввиду неспецифичности исходных условий построения формульного языка утверждать то, что формулы в состоянии описать лексику любого языка мира, приведенную к одному из алфавитов – кириллице или латинице (языки типа армянского, арабского, китайского и т.п.).

Сходное положение в отношении сложных слов было ожидаемо и обнаружено в английском языке, взятом для контроля. Ограничения,

накладываемые на английский язык как аналитический, должны уменьшать число работающих формул, что и отмечается в реальности. Формула 0 – bootjack – jackboot, формула 1 – coastline – sectional, формула 2 – drayman – yardman, формула 3 – phototelegraphy – telephotography, формула 4 – croatian – raincoat, формула 7 – actuator – autocrat, формула 8 – housecat – soutache.

Представляет интерес выяснение того, как усложнение структуры при создании сложных слов влияет на ненаправленное создание изомеров и их формульное представление. Исходя из имеющейся информации, усложнение и высокая структурная специализация живой материи служат причиной того, что происходит вытормаживание ряда процессов за счет ограничения степеней свободы элементов целостной системы. Это касается в первую очередь сокращения числа изомеров в химии и биологии [14, 142-143]. Оказалось, что и языковая изомерия подчиняется этому закону. Обрастание слова морфемами усложняет прежде всего его структуру, его семантику и может менять грамматику, что приводит к тому, что слова с большей длиной (число перестановок теоретически увеличивается) практически не имеют изомеров.

Все это походило бы на формальные упражнения, если бы не вполне системные и внешне совершенно неожиданные нетривиальные результаты, малую часть из которых мы приводим ниже:

1. подобные формулы описывают пары слов двух качественно разных типов – неизомерийную лексику и изомерийную. Таким образом соблюдается необходимая для теории полнота – три десятка формул покрывают все миллионное многообразие словоформ любого языка мира. Изомерия же, как было ранее доказано в цикле работ сотрудников НИЛ теоретической и прикладной лингвистики, поддерживается омонимией, что уже расширяет мощностность формул в другом отношении. При этом лишь изомерийная лексика имеет зеркальные прочтения типа *милом – молим, малом – молам* и т.п. (не следует путать с палиндромами, где одно и то же слово, например, **мадам, потоп, шалаш**, читается слева направо и наоборот);

2. имеются левые и правые формулы. Это понимается так, что в левом столбце таблицы расположены формулы, отвечающие требованию разных начал, тогда как в правом столбце будет лексика с одинаковыми начальными элементами. Тогда левые (четные номера) формулы представляют пары слов с разными началами, правые (нечетные) представляют пары слов с одинаковыми началами. И это значимо – в левом столбце могут быть разнокоренные слова или однокоренные, но с разными префиксами, в правом столбце формулы указывают на однокоренные слова или на разнокоренные слова с одинаковым префиксом;

3. в матрице формул представлена статика и динамика системы. Так, в формуле 4 изомеры *мот – том* имеют возможность развития и реализуют его уже в формуле 6 *мота – тома, мотом – томом* и т.п. Аналогичные переходы от формулы к формуле происходят и при словообразовании;

4. конвергентно-дивергентный стандартный способ удлинения структурной формулы;

5. анализ заполнителей формул подводит к выводу, что с увеличением длины словоформы за счет словообразовательных и словоизменительных морфем специфика языковой материи усложняется в такой степени, что вместо отдельных точек-букв чаще встречаются целые блоки (см. рис. ниже), в которых четко выделяются префиксальные блоки, общая корневая морфема вместе со служебными и флексийные блоки;

6. первичный анализ русской лексики обнаружил первые 20 формул (с 0 по 19 включительно), на болгарском и английском материале выявлены первые 12 (с 0 по 11 включительно). Этих данных уже достаточно для доказательства правомерности формульного представления лексики и даже для выдвижения гипотезы о том, что лексика аналитических языков будет описываться меньшим числом формул в силу менее развитой системы флексий;

7. мощь представлений связана с тем, что внутри каждой формулы благодаря последовательному увеличению длины слов на 1 символ растет число вариантов формулы, где вместо совпадающих/несовпадающих точек отмечаются фрагменты (см. позицию 5) и число вариантов исчисляется с помощью треугольных (1, 3, 6, 10, 15, 21,...), пирамидальных, или тетраэдрических (1, 4, 10, 20, 35,...) и других многоугольных чисел. А эти числа связаны с наиболее плотной упаковкой шаров. Анализ показал, что до длины 8 букв мы имеем дело с полным числом вариантов (плотные упаковки), с увеличением длины слов в формулах плотность снижается, доходя до единственного варианта. Так, для формулы 9 при длине 16 букв теоретически возможно 1365 вариантов, обнаружен пока 1. Далее с увеличением длины изомерийные формулы вырождаются, т.е. становятся ненаполняемыми. Объяснение этой закономерности следующее – языковая материя становится настолько усложненной (обрастание короткого изначально слова префиксами, суффиксами - при словообразовании и присоединение окончаний), что исчезает возможность совпадения как таковая. Указанные особенности вскрывают и основные направления дальнейшей разработки теории языковых упаковок.

Практически настроенные лингвисты сразу готовы спросить – К чему это можно применить? В химии при открытии изомерии как явления существования двух разных веществ с одинаковым составом и разным строением такой вопрос не стоял – одно вещество взрывалось, второе было удобрением. Хотя и там позже было открыто множество разновидностей изомерии – таутомеры, оптические изомеры, цис- и транс-изомеры и т.д. Теоретически настроенный лингвист может осторожно предложить хотя бы два направления.

Первое из них – это проблема записи языковой информации в памяти. Произнесенное и услышанное слово имеет трассу записи. По данным академика П.К.Анохина "извлечение прошлого опыта из памяти происходит по той же нейрхимической трассе, по которой он был зафиксирован в момент приобретения опыта"[1, 91] Естественно, что слова, состоящие из разных звуков, букв, фонем, имеют разные трассы. При каком-либо сходстве это сходство должно отражаться и в трассе записи. Мы не знаем сегодня – как это происходит, положительно зная лишь то, что имеется на входе и что на выходе. Скорость, с которой мы пользуемся огромным числом словоформ при создании текста, огромна. Значит записи должны быть компактными. Этому требованию предлагаемые для обсуждения формулы отвечают.

Так, формула 9 *милом - молим* имеет всего семь точек, представляющих четыре разных символа при трех общих конвергентных точках-узлах. Суммарное число в записи двух слов отдельно – 10 символов. И это без учета того, что длинная запись может читаться частично и это будет в ряде случаев осмысленная правильная словоформа. Так запись *милом – молим* содержит фрагменты **ми, мил, мили, милом, мол, мило, моли, молим, молом**, представляющие правильные словоформы русского языка. При учете этих слов информационная плотность формулы увеличится еще на 25 символов. Всего тогда семь точек, через которые проходит сигнал в состоянии реализовать 35 символов, соответствующих 9 разным словоформам. Если же учесть еще и омонимию – мол (существительное и частица),

моли (от моль и от молить), мило (краткое прилагательное и наречие), то число словоформ возрастет до 12, а число букв, представляющих эти 12 слов до 46. Весьма экономная запись. Предположительно, что более длинные формулы могут иметь значительно большую информационную плотность.

Просматривается и возможность объединения ряда формул в симметрично-асимметричные блоки по 10-15 единиц, превращающие множество формул в сеть. Их информационная мощь еще больше – 23 три точки сети (из них 12 разных) позволяют запись 67 словоформ в сумме представляющих 274 буквы.

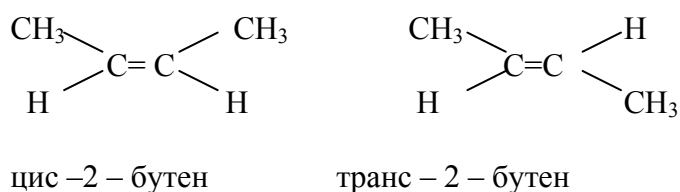
Второе направление связано с обучением иностранным языкам. Овладение вторым языком (близкородственным, допустим, болгарским) – не только увеличивает информационную плотность записи формулы за счет межъязыковой омонимии, но и увеличивает число парных заполнений формул за счет возникновения межъязыковой изомерии – ср. словоформа *галок* (от галка) в русском языке не имеет изомеров, аналогично болгарское *голак* (бедняк, босьяк, оборванец). Знание болгарской лексемы носителем русского языка создает изомерийные отношения *галок* – *голак*, реализованные в формуле 9. Сходная ситуация со словоформой *насут* – *насту* формула 3 и болгарская пара слов *пусат* (уст. оружие) и *пуста* (степь в Венгрии) также формула 3. Знание болгарских лексем создает изомерийную пару *насут* – *пусат* формула 9. Аналогичные примеры можно приводить на болгарском материале и для других формул. Это направление имеет прямой выход в практику – следует сначала для лучшего усвоения давать изомерийную лексику для двух языков, тогда изомер своего родного языка, допустим, "*насут*" будет ассоциироваться со всем изомерийным болгарским гнездом – *пуста*, *пусат*. Внешнее, а по сути своей, системное сходство (симметричное) оболочек слов с необходимостью будет вызывать своеобразную иррадиацию.

Слова, содержащие доминантный комплекс согласных П-С-Т в разных комбинаторных вариантах ПТС, СТП, СПТ, ТСП, ТПС при переменных гласных, создают систему своего рода в виде изомерийного узла.

Следующий этап исследования предполагает анализ формульных изомеров по двум основным направлениям, подразделяющимся на десятки подвидов. Одно из них – общесистемное. Прежде всего, есть основания для обнаружения зеркальной изомерии, цис-транс изомерии, таутомерии и других структурных разновидностей, свойственных химическим и биологическим изомерам. Сам по себе этот факт выводит лингвистику на качественно иной уровень – уровень фундаментальных наук, так как обнаружение явлений, свойственных объектам неживой и живой природы в языке-системе – это доказательство мощи принципов межсистемного изоморфизма.

а. структурный изоморфизм химических и языковых объектов. Сравните структурные формулы двух качественно различных объектов-систем, взятых из двух качественно удаленных друг от друга предметных областей – химии и лингвистики (алкен в химии и глагол в языке):

Рис. 2





Как видно из формул, языковые объекты-глаголы являются своеобразными аналогами цис- и транс- форм химических изомеров. Сами формулы также изоморфны, с той разницей, что мы отдаем себе отчет в том, что в химии это две разные формы существования одного и того же вещества, а в языке – это два вида трасс записи звуковой или графической информации с известным нам семантическим и грамматическим значением, существующие по отдельности в нашем сознании.

б. системный изоморфизм языковых и биологических объектов (текст языковой с пробелами и текст – белковая цепочка без пробелов при заданном алфавите – 20 аминокислот);

в. количественный изоморфизм физики частиц, генетического кода и логики (физика – комбинаторика барионов как долгоживущих частиц из 4-х кварков – u, d, s, c по 3; генетика – комбинаторика из 4-х оснований - А, Г, Ц, У по 3; логика – комбинаторика из 4 типов суждений – общее утверждение, частное утверждение, общее отрицание, частное отрицание по 3) [15].

В лингвистическом плане предельно важным является выяснение всех конвергентных и дивергентных завершений формул, факторы, способствующие двум указанным процессам; частеречный и семантический анализ, анализ фонетических и фонематических изомеров. Важность изучения трех базовых типов изомеров – фонетических, графических и фонематических - видна хотя бы из того факта, что для передачи генетической информации в РНК используются в связи с их функциями и структурными особенностями три класса - рРНК (рибосомальные), тРНК (транспортные) и иРНК/мРНК (информационные, или матричные). Возможно отыскание межсистемного изоморфизма между ними и звуком, буквой и фонемой.

Завершить доклад можно на мажорной оптимистической ноте. Лингвистика может стать, становится и должна стать точной наукой. Для этого она имеет три надежных основания. Это симметрия, асимметрии и диссимметрия в их многочисленных видах и разновидностях.

ЛИТЕРАТУРА

1. Анохин П.К. Философские аспекты теории функциональной системы. Москва, 1978.
2. Урманцев Ю.А. Симметрия природы и природа симметрии. М., Мысль, 1974
3. Карпов В.А. Язык как система. Минск, 1992
4. Барбук С.Г. Анализ лингвистической изомерии на разных уровнях. АКД. Минск, 2000
5. Карпов В.А. Формульное представление лексики // Личность, Слово, Социум / Материалы научно-практической конференции 21-22 февраля 2001, Минск, с.63-67
6. Карпов В.А. Формульное представление изомеров // Актуальные проблемы исследования языка и речи/ Материалы Международной научной конференции молодых ученых, Минск, 30-31 октября 2001, Ч 1. с.46-50
7. Карпов В.А. Проблемы структурной языковой изомерии // Труды ученых лингвистических вузов, Минск, 2001. с.127-133

8. Карпов В.А. Системные методы представления знаний // Материалы IV Международной научной конференции. Ч.1. Минск, БГУ, 2001. с.12-17
9. Карпов В.А. Обоснование симметро-асимметрии в языке // Вестник МГЛУ. серия 1 Филология, № 4; 1998 с.16-22
10. Карпов В.А. Симметрия-асимметрия-диссимметрия // Язык, слово, действительность, Часть 1, Материалы 2 Международной научной конференции, Минск, 2000, с.132-135
11. Карпов В.А. Симметрия и асимметрия как универсальные категории языка // Матэрыялы II Міжнародная канферэнцыі, прысвечанай 75-годзю з дня нараджэння прафесара Л.М.Шакуна, Минск, 2002, Т.2. с.240-247
12. Карпов В.А. Языковые равенства и неравенства // Вестник МГЛУ. серия 1 Филология, № 8; 2001 с.24-32
13. Частотный словарь русского языка. под. ред. Л.Н.Засориной. М.,Русский язык, 1977
14. Ларин Ю.С. Системный подход и эволюционика // Система Симметрия Гармония. М., Мысль, 1988, с.130-144
15. Карпов В.А. Изоморфизм двух систем (междисциплинарный синтез на базе системного подхода) // Веснік БДУ, серыя 3, N3, 1994. с. 26-30