МАТЕМАТИЧЕСКИЕ МОДЕЛИ В ИНФОРМАТИКЕ
(Вычислительные системы)
2002 год Выпуск 169

УДК 510.649

ЛИНГВИСТИЧЕСКИЕ РАЗМЫТЫЕ МОДЕЛИ

Г.Э. Яхъяева

В последнее время наше представление о языке необычайно расширилось. Сейчас изучением языка занимаются не только лингвисты, но и представители других, казалось бы совсем мало связанных между собой, разделов знаний: искусствоведы, биологи, кибернетики, математики и логики.

Легко объяснить глубокий интерес к языку у логиков и математиков. В речевом поведении людей широко используются логические построения. Отсюда, естественно, напрашивается мысль о возможности создания абстрактных лингвистических и логических моделей. Перед математиками возникли и чисто прикладные задачи лингвистического характера — это прежде всего построение искусственных языков программирования для взаимодействия человека с вычислительной машиной и построение алгоритмов машинного перевода текста с одного естественного языка на другой.

Многообразие обыденного языка считается его самым существенным признаком. Это отнюдь не показатель его ущербности. Естественный язык именно в силу его полиморфизма богаче всякого искусственно создаваемого языка.

Концепция вариативности является объяснением этого свойства. Нечеткие и неотчетливые по своему смыслу слова с неровными краями областей их значений, неясность разграничительных линий между понятиями, их многообразие и пестрота являются феноменом естественного языка, отличающего его от любого формального.

Каждое понятие трудно охарактеризовать одним конкретным словом. Для этого, обычно, требуется целый набор слов. Например, если вы попытаетесь поставить в соответствие некоторому понятию слово **считать**, то второму человеку трудно будет определить, какое именно понятие вы имеете в виду. Если же вы выделите пару слов **считать, рассматривать** или пару **считать, подсчитывать**, то вашему собеседнику гораздо легче будет уяснить, о каком понятии идет речь. Необходимо заметить, что разные слова с разной степенью точности отражают то или иное понятие.

Формализуем эту идею. Пусть нам задано некоторое множество слов Ω естественного языка, которое будем считать универсумом. Подмножества этого множества, определяющие данные понятия, задают структуру нечетких множеств, определенных на Ω, т.е. множеств с нечетким отношением принадлежности (см., например, [1]). Весовые характеристики отражения данного понятия каждым словом можно подсчитывать с помощью стохастической взаимосвязи между употребляемостью и многозначностью слова [2]. Каждое такое нечеткое множество будем называть *семантическим полем.* Согласно лингвистическому словарю, семантическим полем называется совокупность слов и выражений языка, составляющих тематический ряд и в своей совокупности покрывающих определенную область значений [3]. Слова, принадлежащие одному семантическому полю, будем называть *семантически подобными.*

Итак, для того, чтобы ваш собеседник смог более точно понять смысл, который вы вкладываете в данное слово, вы должны "подсказать" ему семантическое поле, которому это слово принадлежит. Формализация этой процедуры описывается в § 2.

А как же быть, если вы и ваш собеседник говорите на разных языках? Если даже ваш собеседник знает перевод всех слов вашего словарного запаса на свой язык, вам будет трудно передать смысл, который вы вкладываете в сказанное слово. Ведь помимо того, что ваше слово может принадлежать разным семантическим полям, его перевод также может соответствовать разным понятиям. Для того, чтобы преодолеть эту проблему, на множестве Ω_0 слов вашего естественного языка и на множестве

Ω_1 слов языка вашего собеседника зададим структуры нечетких множеств, поставленных в соответствие одному и тому же набору понятий.

Эта идея полностью согласуется с работой А.Вежбицкой по созданию универсального семантического метаязыка [4]. По предположению А.Вежбицкой семантический метаязык составляется из слов данного языка, которое удовлетворяют следующим условиям:

1) сами по себе эти слова семантически неразложимы, но с их помощью можно осуществлять семантическое разложение других слов данного языка;

2) эти слова имеют переводы на все другие языки, и в каждом из языков совокупность переводов может играть роль семантического метаязыка.

Очевидно, что на множествах Ω_0 и Ω_1 нельзя задать взаимно-однозначное отображение "перевода". Однако, если поставить в соответствие друг другу семантические поля, отвечающие одному понятию, то множества семантических полей будут обладать таким отображением. Теперь, для того, чтобы собеседник вас понял, вместе со словом вы должны указать семантическое поле, которому это слово принадлежит. Тогда, производя перевод всего семантического поля, ваш собеседник безошибочно определит то понятие, которое вы хотели охарактеризовать. Весовые же характеристики слов позволят ему подобрать в своем словаре более удачный перевод вашего слова. Формализация этой процедуры описана в § 3.

Предлагаемые нами модели не претендуют на отражение всех граней естественных языков. Мы просто надеемся, что данные исследования станут еще одним шагом на пути совершенствования формализации работы с естественными языками.

§ 1. Основные определения и обозначения

Теория нечетких множеств, развивающаяся после публикации в 1965 г. работы Л.Заде [1], представляет собой обобщение и переосмысление важнейших направлений классической математики.

Нечеткое множество образуется путем введения обобщенного понятия принадлежности, т.е. расширения двухэлементного множества значений характеристической функции $\{0, 1\}$ до континуума $[0, 1]$. Это означает, что переход от абсолютной принадлежности объекта классу к абсолютной его непринадлежности происходит не скачком, а плавно, постепенно, причем принадлежность элемента множеству выражается числом из интервала $[0, 1]$. Нечеткое множество $A = \{(x, \mu_A(x))\}$ определяется математически как совокупность упорядоченных пар, составленных из элементов x универсального множества X и соответствующих степеней принадлежности $\mu_A(x)$ или непосредственно в виде функции $\mu_A : X \to [0, 1]$.

Нечеткое бинарное отношение можно понимать как нечеткое множество, носителем которого является прямое произведение двух множеств, т.е. $\mu_R : A \times B \to [0, 1]$.

Пусть R — нечеткое бинарное отношение, определенное на прямом произведении $A \times A$. По аналогии с классической математикой, определяются следующие свойства:

- *рефлексивность* $\forall x, y \in A(x \neq y \Rightarrow R(x, y) < R(x, x) = 1)$;
- *симметричность* $\forall x, y \in A(R(x, y) = R(x, y))$;
- *транзитивность* $\forall \alpha \in [0, 1] \forall x, y, z \in A(R(x, y) > \alpha \wedge \wedge R(y, z) > \alpha \Rightarrow R(x, z) > \alpha)$.

Обобщая понятие нечеткого отношения, мы приходим к понятию нечеткой операции [7]. Рассмотрим некоторое отображение $f : A^n \to P(A)$, где $P(A)$ — множество всех подмножеств произвольного множества A.

Отображение f назовем *нечеткой n-арной операцией*, если для каждого $S = f(a_1, \ldots, a_n) \in P(A)$ задана функция приоритета μ_S, определяющая предпочтение или "вес" каждого элемента из S, такая, что для любого $s \in S$ имеем $\mu_S(s) \in (0, 1]$.

Если вне множества S функцию приоритета μ_S считать нулевой, то она фактически задает саму операцию, т.е.

$$\mu(a_1, \ldots, a_n, s) = \begin{cases} \mu_S(s), & s \in S, \\ 0, & s \notin S; \end{cases}$$

$$f(a_1, \ldots, a_n) = \{s | \mu(a_1, \ldots, a_n, s) > 0\} = S.$$

Для бинарной нечеткой операции можно использовать и традиционную запись, введя понятие равенства с данным весом $a * b =_\alpha s$, где $\alpha = \mu(a, b, s)$.

Нечеткую бинарную операцию $*$ назовем *ассоциативной*, если $(a*(b*c) =_\alpha d \wedge (a*b)*c =_\beta d) \Rightarrow (\alpha > 0 \Leftrightarrow \beta > 0)$.

Нечеткую операцию $*$ назовем *компактной*, если для любых $a, b, c, d \in A$ выполняется условие $a * b =_\alpha c \Rightarrow \big((a*b)*d = c*d \wedge \wedge\, d*(a*b) = d*c\big)$.

Элемент $e \in A$ будем называть *нечетко нейтральным*, если для любого $a \in A$ найдутся такие числа $\alpha, \beta \in (0,1]$, что $e * a =_\alpha a; \quad a * e =_\beta a$.

Элемент $b \in A$ будем называть *нечетко обратным* к элементу $a \in A$ относительно нейтрального элемента e и обозначать a^{-e}, если для некоторого $e \in E$ найдутся такие числа $\alpha, \beta \in (0,1]$, что $a * b =_\alpha e, \quad b * a =_\beta e$.

Алгебраическая система $< A, * >$ называется *размытой группой*, если

1) операция ассоциативна,

2) операция компактна,

3) в множестве A существует подмножество E нейтральных элементов,

4) для каждого элемента $a \in A$ существует по крайней мере один нечетко обратный элемент.

В данной статье мы описываем несколько приложений теории размытых групп к лингвистике. В § 2 в качестве универсума берется некоторое конечное множество слов естественного языка. Бинарная операция, определенная на этом множестве, будет интерпретироваться как подбор слов, семантически подобных двум данным. Очевидно, что результат выполнения такой операции будет не однозначным, т.е. размытым. Например, при "умножении" слова **отличный** на слово **непохожий** мы можем получить набор слов

отличный, непохожий, несхожий, несходный.

А при "умножении" того же слова **отличный** на слово **прекрасный**, мы можем получить набор слов

отличный, прекрасный, отменный
превосходный, великолепный, распрекрасный.

В статье показывается, что множество слов естественного языка с такой операцией будет образовывать размытую группу нейтральных элементов, так как каждое слово будет обладать свойствами нечетко нейтрального элемента.

В § 3 в качестве универсума будет рассматриваться объединение множеств слов из двух различных естественных языков. Интерпретация же размытой операции будет разбиваться на две части: 1) подбор семантически подобных слов; 2) понятийный перевод с одного языка на другой. Будет показано, что полученная система образовывает размытую группу, порядок каждого элемента которой будет равен двум. В § 4 будет показан алгоритм построения аналогичной размытой группы для произвольного конечного числа естественных языков.

Любая размытая операция характеризуется тем, что каждый отдельный результат ее выполнения подсчитывается со своим коэффициентом предпочтения. В данной статье для подсчета этого коэффициента мы будем использовать понятие треугольной нормы (сокращенно t–нормы). Изучением t–норм занимаются многие математики, работающие в теории "fuzzy" (см., например, [6]). Дадим определение t–нормы.

Треугольной нормой называется двухместная действительная функция $T : [0,1]^2 \to [0,1]$, удовлетворяющая следующим условиям:

1) $T(0,0) = 0$; $T(\alpha, 1) = T(1, \alpha) = \alpha$;

2) $\alpha_1 \leq \beta_1 \wedge \alpha_2 \leq \beta_2 \Rightarrow T(\alpha_1, \alpha_2) \leq T(\beta_1, \beta_2)$;

3) $T(\alpha, \beta) = T(\beta, \alpha)$;

4) $T(\alpha, T(\beta, \gamma)) = T(T(\alpha, \beta), \gamma)$.

Для наших рассуждений понадобятся t–нормы, обладающие свойством: $f(\delta, \alpha) = f(\alpha, \delta) = \delta$, $\alpha \in [\delta, 1]$. Мы будем называть такие t–нормы *δ–ограниченными t–нормами.*

Можно показать, что существует бесконечно много различных δ–ограниченных t–норм. Для этого достаточно показать, что любую произвольную t–норму можно преобразовать в δ–ограниченную t–норму. Пусть T — некоторая t–норма. Тогда функция

$$f(x,y) = \delta T\Big(1 - \frac{\delta - x}{\delta}; 1 - \frac{\delta - y}{\delta}\Big) + (1-\delta) T\Big(\frac{x-\delta}{1-\delta}; \frac{y-\delta}{1-\delta}\Big)$$

также будет t–нормой, но помимо этого, она будет обладать и свойством δ–ограниченности.

Одним из примеров δ–ограниченной t–нормой является оператор min.

§ 2. Алгоритм определения семантического поля данного слова

Пусть нам дано конечное множество $\Omega = \{\omega_1, \dots, \omega_n\}$ слов некоторого естественного языка. Вопрос о том, является ли конечным или счетным множество всех слов данного естественного языка — спорный вопрос. Однако в любой реальной системе может быть использовано лишь конечное число слов. Опираясь на эти соображения далее всегда будем считать, что множество Ω конечно.

Допустим, что на множестве Ω задано бинарное нечеткое отношение S, отражающее степень принадлежности двух слов одному семантическому полю. Выберем достаточно малое $\delta \in (0, 1)$ и будем считать, что $S(\omega_i, \omega_j) = \delta$ в том и только в том случае, когда слова ω_i и ω_j лежат в разных семантических полях. Таким образом, отношение S является связанным отношением, т.е. $\forall \omega_i, \omega_j \in \Omega(S(\omega_i, \omega_j) \geq \delta > 0)$. Это можно проинтерпретировать тем, что степень δ характеризует тот факт, что слова взяты из одного естественного языка. Далее будет показано, что, варьируя значение параметра δ, мы можем изменять "глубину" раскрытия семантического поля.

Нечеткое отношение S должно удовлетворять условию рефлексивности, т.е. $\forall \omega_i, \omega_j \in \Omega(\omega_i \neq \omega_j \Rightarrow S(\omega_i, \omega_j) < S(\omega_i, \omega_i) = 1)$. В самом деле, насколько близки по смыслу не были бы два слова, всегда можно найти нюансы, отличающие их друг от друга. Любое слово "на сто процентов" семантически подобно только самому себе.

Не вызывает сомнение и тот факт, что данное отношение должно быть симметричным, т.е. $\forall \omega_i, \omega_j \in \Omega(S(\omega_i, \omega_j) = S(\omega_j, \omega_i))$.

Было бы замечательно, если бы отношение S обладало бы свойством транзитивности. Но в любом естественном языке существует огромное количество омонимов, т.е. слов, которые принадлежат сразу нескольким семантическим полям. Так, согласно словарю [5], слово **хозяин** принадлежит следующим семантическим полям:

хозяин – повелитель – владыка – властелин – господин,
хозяин – владелец – собственник – обладатель.

Тогда мы получим

$$S(\textbf{владыка, хозяин}) > \delta \wedge S(\textbf{хозяин, собственник}) > \delta,$$

Однако

$$S(\textbf{владыка, собственник}) = \delta.$$

Именно этот факт является слабым звеном многих программ подбора синонимов. Переходя от одного синонима к другому, система постепенно забывает исходную информацию, что может привести к довольно абсурдным подборкам. Так, например, работает система подбора синонимов, встроенная в общеизвестный текстовый редактор Word. Если вы попытаетесь подобрать синоним к слову **содержит**, система предложит вам четыре различных слова: **держит, хранит, заключает, кормит**, которые не являются синонимами друг друга. Однако каждое из них, в паре со словом **содержит**, может иметь еще несколько синонимов. Но если продолжить процедуру, то система будет подбирать синоним для выделенного слова, забыв об исходном. Покажем, к чему это может привести:

содержит – держит – сдерживает – тормозит;
содержит – хранит – бережет – экономит;
содержит – заключает – охватывает – обнимает;
содержит – кормит – питает – впитывает – всасывает.

Итак, в качестве исходных данных, мы имеем нечеткое бинарное отношение, обладающее свойствами рефлексивности и симметричности. В силу конечности универсума, это отношение можно задать матрицей. В табл. 1 приведен пример такой матрицы для множества Ω, состоящего из 24 слов. Единицы в этой матрице стоят только по главной диагонали, а сама матрица является симметричной, что и обеспечивает выполнение вышеперечисленных свойств.

На множестве Ω определим размытую операцию, которая будет интерпретировать действие подбора семантически подобных слов. Выполняя очередное действие, мы будем получать семантически подобные слова для всех тех слов, которые были выбраны в результате совершения предыдущих действий.

Пусть f — некоторая δ–ограниченная t–норма. Размытая операция $*$ будет задаваться следующим рекурсивным правилом:

1) для любых $\omega_i, \omega_j, \omega_k \in \Omega$

$$\omega_i * \omega_j =_{f(S(\omega_i, \omega_k), S(\omega_j, \omega_k))} \omega_k;$$

2) пусть

$$\omega_{i_1} * \omega_{i_2} * \ldots * \omega_{i_n} =_{\alpha_1} \omega_k, \quad \omega_{j_1} * \omega_{j_2} * \ldots * \omega_{j_m} =_{\alpha_2} \omega_k,$$

тогда

$$(\omega_{i_1} * \ldots * \omega_{i_n}) * (\omega_{j_1} * \ldots * \omega_{j_m}) =_{f(\alpha_1, \alpha_2)} \omega_k.$$

ПРИМЕР 1. Рассмотрим множество Ω и отношение S, заданные табл.1. Пусть $\delta = 0,1$; $f(x,y) = \min(x,y)$. Мы имеем

$$\textbf{тишина}*\textbf{безмолвие} =_{\min(0.8;0.8)} \textbf{спокойствие},$$

и это означает, что слово **спокойствие** с приоритетом 0,8 принадлежит семантическому полю, образованному словами **тишина** и **безмолвие**.

С другой стороны, мы имеем

$$\textbf{тишина} * \textbf{безмолвие} =_{\min(0.3;\delta)} \textbf{мир},$$

и это означает, что слово **мир** не принадлежит семантическому полю, образованному словами **тишина** и **безмолвие**.

Произведение

$$\textbf{гармония} * \textbf{слой}$$

не дают ни одного слова с приоритетом больше δ. Следовательно, в Ω не существует семантического поля, содержащего эти два слова.

Заметим, что объективность данных рассуждений всецело зависит от объективности составления табл.1. На данном этапе исследований мы не ставили себе задачу составления такой таблицы, поэтому, будем считать, что она отражает семантику естественного языка в некотором заданном контексте.

В силу связанности отношения S, автоматически выполняются все аксиомы размытой группы [7]. Следовательно, алгебраическая система $< \Omega, * >$ изоморфна некоторой размытой группе нейтральных элементов [8].

Согласно свойствам t–нормы, данная система обладает свойствами сильной коммутативности и сильной ассоциативности, т.е.

$$\forall \omega_i, \omega_j, \omega_k[(\omega_i * \omega_j =_\alpha \omega_k \wedge \omega_j * \omega_i =_\beta \omega_k) \Rightarrow \alpha = \beta],$$

$$\forall \omega_i, \omega_j, \omega_k, \omega_l[((\omega_i * \omega_j) * \omega_k =_\alpha \omega_l \wedge \omega_i * (\omega_j * \omega_k) =_\beta \omega_l) \Rightarrow \alpha = \beta].$$

Это хорошо согласуется с интерпретацией операции как действия подбора общих семантически подобных слов. Более того, сильная ассоциативность позволит нам в дальнейшем опускать скобки, когда будет рассматриваться конечное произведение элементов из Ω.

Очевидно, что $\omega_{i_1} * \ldots * \omega_{i_n} = \Omega$. Под *усиленным произведением* элементов $\omega_{i_1}, \ldots, \omega_{i_n}$ будем понимать множество

$$\overline{\omega_{i_1} * \ldots * \omega_{i_n}} = \{\omega | \omega_{i_1} * \ldots * \omega_{i_n} =_\alpha \omega \wedge \alpha > \delta\}.$$

Допустим, что $\overline{\omega_{i_1} * \ldots * \omega_{i_n}} = \{\pi_1, \ldots, \pi_k\}$. Тогда определим множество

$$\overline{\overline{\omega_{i_1} * \ldots * \omega_{i_n}}} = \overline{\pi_1 * \ldots * \pi_k}.$$

Далее, пусть нам даны слова $\omega^1, \omega^2, \ldots, \omega^n \in \Omega$, а также некоторое подмножество слов $\Omega' \subset \Omega$, не содержащее этих слов. Под *семантически самым близким словом* для слов $\omega^1, \omega^2, \ldots, \omega^n$ из набора Ω' будем считать слово ω', отвечающее следующему условию: $\omega^1 * \ldots * \omega^n =_\alpha \omega'$ и $\alpha = \max\{\beta | \omega^1 * \ldots \ldots * \omega^n =_\beta \omega_i,\ \omega_i \in \Omega'\}$.

Опишем теперь алгоритм определения семантического поля для заданного слова $\omega \in \Omega$. Конечное число раз наш алгоритм может обращаться к оракулу за дополнительной информацией. Заметим, что в качестве оракула может выступать как эксперт, так и некая статистическая машина.

Шаг 1. Вычисляется усиленное произведение $A^* = \overline{\omega * \omega}$, которое дает нам все семантически подобные слова для заданного слова. Дальнейшая наша задача из полученного множества выделить одно конкретное семантическое поле.

Заметим, что множество $\overline{\omega * \omega}$ не может быть пустым, по крайней мере один элемент ω оно содержит. Если оно содержит ровно один элемент, то слово ω не имеет семантически подобных слов, и алгоритм на этом заканчивает свою работу.

Если $\overline{\omega * \omega} = \overline{\overline{\omega * \omega}}$, то слову ω соответствует единственное семантическое поле, состоящее из всех слов множества A^*.

В противном случае мы можем вычеркнуть из множества A^* слова, принадлежащие множеству $\overline{\overline{\omega * \omega}}$, так как они несут столько же информации, как и слово ω и будут заставлять алгоритм работать вхолостую.

Итак, мы имеем $A^* = \overline{\omega * \omega} - \overline{\overline{\omega * \omega}}$. Определим $A = A^*$ и $a^1 = \omega$.

Шаг $2k$. В множестве A находим семантически самое близкое слово a к слову ω и вычисляем

$$\overline{a^1 * \ldots * a^k * a}. \tag{1}$$

Элемент a записываем в множество B_k, а из множества A вычеркиваем все слова, принадлежащие усиленному произведению (1). После чего в множестве A снова находим семантически самое близкое слово к слову ω и повторяем процедуру до тех пор, пока множество A не станет пустым.

Шаг $2k + 1$. Обращаемся к оракулу с просьбой выбрать одно слово из множества B_k. Заметим, если множество B_k состоит только из одного элемента, то нет необходимости обращаться к оракулу. Этот элемент выбирается автоматически.

Допустим, что оракул выбрал слово b. Тогда вычисляем произведения

$$\overline{a^1 * \ldots * a^k * b} \tag{2}$$

и

$$\overline{\overline{a^1 * \ldots * a^k * b}}. \tag{3}$$

Если множества (2) и (3) равны, то алгоритм заканчивает свою работу, выдав в качестве ответа все слова, принадлежащие этим множествам.

Т а б л и ц а 1

Нечеткое отношение семантического подобия, $\delta = 0,1$ (для большей наглядности в таблице δ обозначено прочерком)

№ п/п	Слово	1 Безмолвие	2 Безмятежность	3 Вселенная	4 Гармония	5 Единство	6 Земля	7 Космос	8 Круг	9 Мир	10 Мироздание	11 Область	12 Общество	13 Общность	14 Окружение	15 Планета	16 Покой	17 Свет	18 Слой	19 Согласие	20 Спокойствие	21 Среда	22 Сфера	23 Тишина	24 Цельность
1	Безмолвие	1	0,6	–	–	–	–	–	–	–	–	–	–	–	–	–	0,8	–	–		0,8	–	–	0.9	–
2	Безмятежность	0,6	1	–	–	–	–	–	–	–	–	–	–	–	–	–	0,9	–	–	–	0,9	–	–	0,7	–
3	Вселенная	–	–	1	0,2	0,2	0,4	0,9	–	0,9	0,8		0,2		–	0,2		0,4	–	–	–	–	–	–	0,2
4	Гармония	–	–	0,2	1	0,7	–	–	–	0,8	0,3	–	–	0,7	–	–	–	–	–	0,9	–	–	–	–	0,7
5	Единство	–	–	0,2	0,7	1	–	–	–	0,6	0,2	–	–	0,8	–	–	–	–	–	0,8	–	–	–	–	0,9
6	Земля	–	–	0,4	–	–	1	0,3	–	0,9	0,2	–	–	–	0,4	–	–	1	0,3	-0,9	0,2	–	–	–	–
7	Космос	–	–	0,9	–	–	0,3	1	–	0,9	0,8	–	0,2	–	–	0,3	–	0,7	–	–	–	–	–	–	–
8	Круг	–	–	–	–	–	–	–	1	0,5	–	0,8	0,3	0,2	0,9	–	–	–	0,7	–	–	0,3	0,9	–	–
9	Мир	–	–	0,9	0,8	0,6	0,9	0,9	0,5	1	0,7	0,4	0,5	0,9	0,2	0,9	0,9	0,6	–	0,8	0,8	0,6	0,4	0,5	0,5
10	Мироздание	–		0,8	0,3	0,2	0,2	0,8	–	0,7	1	–	–	–	–	0,3		0,8	–			–	–	–	0,2
11	Область	–	–	–	–	–	–	–	0,8	0,4	–	1	0,2	–	0,8	–	–	–	–	–	–	0,2	0,9	–	–
12	Общество	–	–	0,2	–	–	–	0,2	0,3	0,5	–	0,2	1	0,6	0,2	–	–	0,7	0,6	–	–	0,5	0,2	–	–
13	Общность	–	–	–	0,7	0,8	–	–	0,2	0,9	–	–	0,6	1	–	–	–	–	–	–	0,9	–	0,7	0,4	–
14	Окружение	–	–	–	–	–	–	–	0,9	0,2	–	0,8	0,2	–	1	–	–	–	0,6	–	–	0,5	0,2	–	–
15	Планета	–	–	0,2	–	–	0,9	0,3	–	0,9	0,3	–	–	–	–	1	–	0,9	–		–	0,2	0,2	–	–
16	Покой	0,8	0,9	–	–	–	–	–	–	0,9	–	–	–	–	–	–	1	–	–	–	0,9	–	–	0,7	–
17	Свет	–	–	0,4	–	–	0,9	0,7	–	0,6	0,8	–	0,7	–	–	0,9	–	1	–	–	–	–	–	–	–
18	Слой	–	–	–	–	–	–	–	0,7	–	–	–	0,6	–	0,6	–	–	–	1	–	–	0,7	–	–	–
19	Согласие	–	–	–	0,9	0,8	–	–	–	0,8	–	–	–	0,9	–	–	–	–	–	1	–	–	–	–	0,6
20	Спокойствие	0,8	0,9	–	–	–	–	–	–	0,8	–	–	–	–	–	–	0,9	–	–	–	1	–	–	0,8	–
21	Среда	–	–	–	–	–	0,2	–	0,3	0,6	–	0,2	0,5	0,7	0,5	0,2	–	–	0,7	–	–	1	0,2	–	–
22	Сфера	–	–	–	–	–	0,2	–	0,9	0,4	–	0,9	0,2	0,4	0,2	0,2	–	–	–	–	–	0,2	1	–	–
23	Тишина	0,9	0,7	–	–	–	–	–	–	0,3	–	–	–	–	–	–	0,7	–	–	–	0,8	–	–	1	–
24	Цельность	–	–	0,2	0,7	0,9	–	–	–	0,5	0,2	–	–	–	–	–	–	–	–	0,6	–	–	–	–	1

Т а б л и ц а 2

Матрица слов, семантически подобных слову мир

№ п/п	Слово	3 Вселенная	4 Гармония	5 Единство	6 Земля	7 Космос	8 Круг	9 Мир	10 Мироздание	11 Область	12 Общество	13 Общность	14 Окружение	15 Планета	16 Покой	17 Свет	19 Согласие	20 Спокойствие	210 Среда	22 Сфера	23 Тишина	24 Цельность
3	**Вселенная**	1	0,2	0,2	0,4	0,9	–	**0,9**	0,8	–	0,2	–	–	0,2	–	0,4	–	–	–	–	–	0,2
4	**Гармония**	0,2	1	0,7	–	–	–	**0,8**	0,3	–	–	0,7	–	–	–	–	0,9	–	–	–	–	0,7
5	**Единство**	0,2	0,7	1	–	–	–	**0,6**	0,2	–	–	0,8	–	–	–	–	0,8	–			–	0,9
6	**Земля**	0,4	–	–	1	0,3	–	**0,9**	0,2	–	–	–	–	0,9	–	0,9	–	–	0,2	0,2	–	–
7	**Космос**	0,9	–	–	0,3	1	–	**0,9**	0,8	–	0,2	–	–	0,3		0,7	–	–	–	–	–	–
8	**Круг**	–	–	–	–	–	1	**0,5**	–	0,8	0,3	0,2	0,9	–	–	–	–	–	0,3	0,9	–	–
9	**Мир**	**0,9**	**0,8**	**0,6**	**0,9**	**0,9**	**0,5**	**1**	**0,7**	**0,4**	**0,5**	**0,9**	**0,2**	**0,9**	**0,9**	**0,6**	**0,8**	**0,8**	**0,6**	**0,4**	**0,3**	**0,5**
10	**Мироздание**	0,8	0,3	0,2	0,2	0,8	–	**0,7**	1	–	–	–	–	0,3	–	0,8	–	–	–	–	–	0,2
11	**Область**	–	–	–	–	–	0,8	**0,4**	–	1	0,2	–	0,8	–	–	–	–	–	0,2	0,9	–	–
12	**Общество**	0,2	–	–	–	0,2	0,3	**0,5**	–	0,2	1	0,6	0,2	–	–	0,7	–	–	0,5	0,2	–	–
13	**Общность**	–	0,7	0,8	–	–	0,2	**0,9**	–	–	0,6	1	–	–	–	–	0,9	–	0,7	0,4	–	–
14	**Окружение**	–	–	–	–	–	0,9	**0,2**	–	0,8	0,2	–	1	–	–	–	–	–	0,5	0,2	–	–
15	**Планета**	0,2	–	–	0,9	0,3	–	**о,9**	0,3	–	–	–	–	1	–	0,9	–	–	0,2	0,2	–	–
16	**Покой**	–	–	–	–	–	–	**0,9**	–	–	–	–	–	–	1	–	–	0,9	–	–	0,7	–
17	**Свет**	0,4	–	–	0,9	0,7	–	**0,6**	0,8	–	0,7	–	–	0,9	–	1	–	–	–	–	–	–
19	**Согласие**	–	0,9	0,8	–	–	–	**0,8**	–	–	–	0,9	–	–	–	–	1	–	–	–	–	0,6
20	**Спокойствие**	–	–	–	–	–	–	**0,8**	–	–	–	–	–	–	0,9	–	–	1	–	–	0,8	–
21	**Среда**	–	–	–	0,2	–	0,3	**0,6**	–	0,2	0,5	0,7	0,5	0,2	–	–	–	–	1	0,2	–	–
22	**Сфера**	–	–	–	0,2	–	0,9	**0,4**	–	0,9	0,2	0,4	0,2	0,2	–	–	–	–	0,2	1	–	–
23	**Тишина**	–	–	–	–	–	–	**0,3**	–	–	–	–	–	–	0,7	–	–	0,8	–	–	1	–
24	**Цельность**	0,2	0,7	0,9	–	–	–	**0,5**	0,2	–	–	–	–	–	–	–	0,6	–	–	–	–	1

В противном случае множество (3) содержится в множестве (2). Тогда конструируем множество A следующим образом $A = (A^* \cap \overline{a^1 * \ldots * a^k * b}) - \overline{\overline{a^1 * \ldots * a^k * b}}$, и определяем $a^{k+1} = b$. После чего переходим к выполнению следующего шага.

В силу того, что множество Ω является конечным, данный алгоритм всегда будет завершать свою работу за конечное число шагов.

ПРИМЕР 2. Пусть множество Ω и отношение S заданы табл.1. Нам нужно найти некоторое семантическое поле слова **мир**.

Шаг 1. Из табл.1 вычеркиваются все строки и столбцы, соответствующие словам, не являющимися семантически подобными слову **мир**. Получится табл. 2, в которой в строке и столбце, соответствующим слову **мир**, нет прочерков. Полученную табл.2 будем обозначать A^*. В нашем примере пришлось вычеркнуть только три строки и три столбца. Однако в системах достаточно большой мощности эта процедура будет значительно сокращать матрицу, а значит и существенно увеличивать скорость работы алгоритма.

Шаг 2. Алгоритм находит семантически самое близкое слово к слову **мир**. Если их оказывается несколько, то фиксируется первый из найденных. Тогда из табл.2 вычеркиваются строки и столбцы, соответствующие словам, семантически подобным зафиксированному слову. Среди невычеркнутых слов, снова ищется самый близкий синоним к слову **мир**. Вся эта процедура повторяется до тех пор, пока не вычеркнутся все слова.

Шаг 3. Алгоритм предлагает оракулу список всех зафиксированных слов с просьбой выбрать одно из них. В нашем примере список выделенных слов будет следующий: **вселенная**, **общность**, **покой**, **круг**. Это означает, что при первом приближении алгоритм определил четыре различных семантических поля слова **мир**.

Допустим оракул выбрал слово **вселенная**. Тогда из таблицы A^* (табл.2) вычеркиваются строки и столбцы, соответствующие словам, которые не являются семантически подобными слову **вселенная**. Строки (и столбцы), соответствующие словам

Т а б л и ц а 3

Матрица слов, семантически подобных словам мир, вселенная

№ п/п	Слово	3	4	5	6	7	10	12	15	17	24
		Мир, вселенная	Гармония	Единство	Земля	Космос	Мироздание	Общество	Планета	Свет	Цельность
3	**Мир, вселенная**	**0,9**	**0,2**	**0,2**	**0,4**	**0,9**	**0,7**	**0,2**	**0,2**	**0,4**	**0,2**
4	Гармония	**0,2**	1	0,7	–	–	0,3	–	–	–	0,7
5	Единство	**0,2**	0,7	1	–	–	0,2	–	–	–	0,9
6	Земля	**0,4**	–	–	1	0,3	0,2	–	0,9	0,9	–
7	Космос	**0,9**	–	–	0,3	1	0,8	0,2	0,3	0,7	–
10	Мироздание	**0,7**	0,3	0,2	0,2	0,8	1	–	0,3	0,8	0,2
12	Общество	**0,2**	–	–	–	0,2	–	1	–	0,7	–
15	Планета	**0,2**	–	–	0,9	0,3	0,3	–	1	0,9	–
17	Свет	**0,4**	–	–	0,9	0,7	0,8	0,7	0,9	1	–
24	Цельность	**0,2**	0,7	0,9	–	–	0,2	–	–	–	1

мир и **вселенная** сливаются с помощью оператора min (см. табл. 3).

Таблица 4

Матрица слов, семантически подобных словам
мир, вселенная, гармония

№ п/п	Слово	3	5	10	24
		Мир, вселенная, гармония	Единство	Мироздание	Цельность
3	**Мир, вселенная, гармония**	**0,2**	**0,2**	**0,3**	**0,2**
5	Единство	**0,2**	1	0,2	0,9
10	Мироздание	**0,3**	0,2	1	0,2
24	Цельность	**0,2**	0,9	0,2	1

Таблица 5

Матрица слов, семантически подобных словам
мир, вселенная, космос

№ п/п	Слово	3	6	10	12	15	17
		Мир, вселенная, космос	Земля	Мироздание	Общество	Планета	Свет
3	**Мир,вселенная, космос**	**0,9**	**0,3**	**0,7**	**0,2**	**0,2**	**0,4**
6	Земля	**0,3**	1	0,2	–	0,9	0,9
10	Мироздание	**0,7**	0,2	1	–	0,3	0,8
12	Общество	**0,2**	–	–	1	–	0,7
15	Планета	**0,2**	0,9	0,3	–	1	0,9
17	Свет	**0,4**	0,9	0,8	0,7	0,9	1

Шаг 4. Алгоритм работает по аналогии с шагом 2, только на этот раз ищутся семантически самые близкие слова к словам **мир и вселенная**.

Шаг 5. Оракулу предоставляется список слов: **космос, гармония**. Если оракул выбирает слово **гармония**, то таблица A^* (табл. 2) принимает вид табл. 4. В табл. 4 нет прочерков, следовательно, алгоритм заканчивает свою работу, выдавая в качестве ответа семантическое поле:

мир, вселенная, гармония, единство, мироздание, цельность.

Если же оракул выбрал слово **космос** (см. табл. 5), то алгоритм вынужден продолжить работу. Он завершит вычисления на седьмом шаге, выдав либо поле:

мир, вселенная, космос, мироздание, земля, планета, свет,

либо

мир, вселенная, космос, общество.

Заметим, что если параметр δ задать более общим, то каждый раз будет вычеркиваться большее количество строк и

Таблица 6

Матрица слов, семантически подобных словам **мир, вселенная**, $\delta = 0.2$

№ п/п	Слово	3 Мир, вселенная,	6 Земля	7 Космос	10 Мироздание	17 Свет
3	**Мир,вселенная,**	**0,9**	0,4	0,9	0,7	0,4
6	**Земля**	**0,4**	1	0,3	0,2	0,9
7	**Космос**	**0,9**	**0,3**	1	0,8	0,9
10	**Мироздание**	**0,7**	0,2	0,8	1	0,8
17	**Свет**	**0,4**	0,9	0,7	0,8	1

столбцов, а семантические поля будут отражать более обобщенный смысл. Так, например, при $\delta = 0,2$, на третьем шаге нашего примера мы получим табл.6, отражающую семантическое поле

мир, вселенная, земля, космос, мироздание, свет.

§ 3. Двуязычный словарь–переводчик

Пусть теперь нам даны два множества $\Omega_0 = \{\omega_1^0, \dots, \omega_N^0\}$ и $\Omega_1 = \{\omega_1^1, \dots, \omega_M^1\}$ слов из двух различных естественных языков. Язык, которому принадлежат слова множества Ω_0, будем называть "родным" языком. Предполагается, что на Ω_0 определено отношение S, и, следовательно, работает алгоритм определения семантического поля. Язык, содержащий слова множества Ω_1, будем называть "иностранным", для которого алгоритм определения семантического поля не определен.

Предполагается, что задано нечеткое бинарное отношение $T \subset \Omega_0 \times \Omega_1$, интерпретирующее перевод слов родного языка на иностранный язык. Запись $T(\omega_i^0, \omega_j^1) = \alpha$, $\alpha \in (\delta, 1]$, будет означать, что иностранное слово ω_j^1 является переводом родного языка ω_i^0 с приоритетностью α. Так же, как и в случае с отношением S, условимся, что если слово ω_j^1 не является переводом слова ω_i^0 считать $T(\omega_i^0, \omega_j^1) = \delta$. Таким образом, мы обеспечили выполнение свойства $\forall \omega_i^0 \in \Omega_0 \forall \omega_j^1 \in \Omega_1 (T(\omega_i^0, \omega_j^1) \geq \delta > 0)$.

Отметим, что отношение T должно обладать следующим свойством. Если слова $\omega_1^0, \dots, \omega_k^0$ принадлежат одному семантическому полю, то они должны иметь общий перевод. Забегая вперед, дадим формальное описание этого свойства

$$\overline{\omega_1^0 * \dots * \omega_k^0} = \{\omega_1^0, \dots, \omega_k^0\} \Rightarrow$$
$$\Rightarrow \exists\, \omega^1 \in \Omega_1 (T(\omega_1^0, \omega^1) > \delta \wedge \dots \wedge T(\omega_k^0, \omega^1) > \delta). \qquad (4)$$

Несмотря на то, что два семантически подобных слова должны обладать общим переводом, нельзя требовать того, чтобы множество переводов одного из них совпадало с множеством переводов другого. Это объясняется тем, что одно из этих слов может оказаться омонимом родного языка.

Необходимо также заметить, что если два или более слов имеют общий перевод, они не обязательно должны быть семантически подобными. Например, слова **прекрасный** и **штраф** в английском языке обозначаются словом **fine**, но, очевидно, они не являются семантически подобными.

На множестве $\Omega = \Omega_0 \cup \Omega_1$ зададим размытую операцию $*$ следующим образом. Для любых слов $\omega_p^{\varepsilon_0}, \omega_{i_1}^{\varepsilon_1}, \ldots, \omega_{i_k}^{\varepsilon_k}$, $\varepsilon_j \in \{0, 1\}$, из Ω определим

$$\omega_{i_1}^{\varepsilon_1} * \omega_{i_2}^{\varepsilon_1} * \ldots * \omega_{i_k}^{\varepsilon_k} = \alpha_p \ \omega_p^{\varepsilon_0}, \tag{5}$$

где $\alpha_p = 0 \Leftrightarrow \varepsilon_p \neq \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_k (mod\ 2)$.

Нетрудно убедиться, что алгебраическая система $< \Omega, * >$ является размытой группой со скелетом порядка 2.

Формула (5) отражает структуру системы $< \Omega, * >$, но не дает никакой информации о весовых значениях операции $*$. В связи с этим мы должны дополнить определение вводимой нами операции. Параллельно с этим мы будем объяснять интерпретацию, вкладываемую нами в данную размытую операцию. Для этого зададим t–норму $f : [\delta, 1] \times [\delta, 1] \to [\delta, 1]$, отвечающую свойству δ–ограниченности. Также для упрощения записи, будем полагать

$$f^n(x_1, x_2, \ldots, x_n) = f(x_1, f(x_2, \ldots, f(x_{n-2}, f(x_{n-1}, x_n)) \ldots)).$$

Допустим, что мы уже совершили $n - 1$ действий, в результате выполнения которых мы получили все элементы множества Ω_0, либо множества Ω_1 (каждый со своим приоритетом). Интерпретацию выполнения n–го действия можно подразделить на два случая.

Случай 1. *Умножение на элемент из* Ω_0. При выполнении этого действия мы будем получать набор семантически подобных слов в том языке, в котором находились до совершения действия.

Случай 1.1. Если мы находились в родном языке, то эта операция идентична той, что описана в предыдущем пункте. Пусть

$$\omega_{i_1}^{\varepsilon_1} * \ldots * \omega_{i_n}^{\varepsilon_n} = \alpha_j \ \omega_j^0, \ \ j = 1, \ldots, N.$$

Тогда
$$(\omega_{i_1}^{\varepsilon_1} * \ldots * \omega_{i_n}^{\varepsilon_n}) * \omega_p^0 =_{f(\alpha_j, S(\omega_p^0, \omega_j^0))} \omega_j^0.$$

Случай 1.2. Если же мы находились в множестве иностранных слов, то производя умножение на "родное" слово ω_p^0, мы получим набор семантически подобных слов для заданных иностранных слов, которые в то же время являются переводом выбранного родного слова. Это действие возможно задать с помощью следующей процедуры. Сперва проверяются, является ли слово ω_p^0 переводом заданных иностранных слов. Если проверка проходит успешно, в множестве Ω_1 подбираются все переводы слова ω_p^0. Весовые значения операции вычисляются с учетом всех использованных в процедуре параметров.

Дадим формальное описание этой процедуры. Пусть
$$\omega_{i_1}^{\varepsilon_1} * \ldots * \omega_{i_n}^{\varepsilon_n} =_{\alpha_j} \omega_j^0, \quad j = 1, \ldots, M.$$

Из множества весовых значений $\{\alpha_j | j = 1, \ldots, M\}$ выберем подмножество $\{\alpha_{j_1}, \ldots, \alpha_{j_k}\}$ всех весовых значений больших δ. Множество соответствующих им слов $\{\omega_{j_1}^1, \ldots, \omega_{j_k}^1\}$ является тем множеством иностранных слов, для которых в ходе операции необходимо подобрать семантически подобные слова. Тогда $(\omega_{i_1}^{\varepsilon_1} * \ldots * \omega_{i_n}^{\varepsilon_n}) * \omega_p^0 =_\gamma \omega_r^1$, где $\gamma = f^{k+1}(f(\alpha_{j_1}, T(\omega_p^0, \omega_{j_1}^1)), \ldots$ $\ldots, f(\alpha_{j_k}, T(\omega_p^0, \omega_{j_k}^1)), T(\omega_p^0, \omega_r^1))$.

Случай 2. *Умножение на элемент из* Ω_1. При выполнении этого действия мы будем получать перевод заданных слов на другой язык, с поправкой на выбранное иностранное слово.

Случай 2.1. Если мы находимся в родном языке, то нашу операцию можно интерпретировать как пересечение переводов заданных "родных" слов на иностранный язык в случае, если выбранное иностранное слово ω_p^1 является переводом всех этих слов с приоритетностью больше δ.

Итак, мы имеем $\omega_{i_1}^{\varepsilon_1} * \ldots * \omega_{i_n}^{\varepsilon_n} =_{\alpha_j} \omega_j^0, \quad j = 1, \ldots, N$. Так же, как и в предыдущем случае, выбираем множество весовых значений $\{\alpha_{j_1}, \ldots, \alpha_{j_k}\}$ больше δ. И вычисляем
$$(\omega_{i_1}^{\varepsilon_1} * \ldots * \omega_{i_n}^{\varepsilon_n}) * \omega_p^1 =_\gamma \omega_r^1,$$

где $\gamma = f^k(f^2(\alpha_{j_1}, T(\omega^0_{j_1}, \omega^1_p), T(\omega^0_{j_1}, \omega^1_r)), \ldots$
$$\ldots, f^2(\alpha_{j_k}, T(\omega^0_{j_k}, \omega^1_p), T(\omega^0_{j_k}, \omega^1_r))).$$

Случай 2.2. Если же мы находимся в иностранном языке, то ищется совместный перевод заданных и выбранных иностранных слов на родной язык. Пусть $\omega^{\varepsilon_1}_{i_1} * \ldots * \omega^{\varepsilon_n}_{i_n} =_{\alpha_j} \omega^0_j$, $j = 1, \ldots, M$. Снова выбираем множество весовых значений $\{\alpha_{j_1}, \ldots, \alpha_{j_k}\}$ больше δ и вычисляем

$$(\omega^{\varepsilon_1}_{i_1} * \ldots * \omega^{\varepsilon_n}_{i_n}) * \omega^1_p =_\gamma \omega^0_r,$$

где $\gamma = f^{k+1}(f(\alpha_{j_1}, T(\omega^0_r, \omega^1_{j_1})), \ldots$
$$\ldots, f(\alpha_{j_k}, T(\omega^0_r, \omega^1_{j_k})), T(\omega^0_r, \omega^1_p))).$$

Замечание 1. Во всех описанных процедурах в случае $n = 1$ условимся считать $\omega^{\varepsilon_1}_{i_1} =_1 \omega^{\varepsilon_1}_{i_1}$ и $\omega^{\varepsilon_1}_{i_1} =_\delta \omega^{\varepsilon_1}_j$ для всех $j \neq i_1$.

Замечание 2. Описанная выше операция $*$ обладает свойством слабой ассоциативности, и даже усиленным свойством ассоциативности, т.е. для любых $\omega_1, \omega_2, \omega_3, \omega_4 \in \Omega$ имеем

$$((\omega_1 * \omega_2) * \omega_3 =_\alpha \omega_4 \wedge \omega_1 * (\omega_2 * \omega_3) =_\beta \omega_4) \Rightarrow (\alpha > \delta \Leftrightarrow \beta > \delta).$$

Однако свойство сильной ассоциативности, в общем случае, может и не выполняться. Но, так как в данной интерпретации мы не нуждаемся в иной расстановке скобок, кроме той, что рассмотрена выше, мы условимся приравнивать весовые значения конечного произведения элементов из Ω при любом ином способе расстановки скобок значениям, полученным при данной расстановке скобок.

Теперь опишем два алгоритма, работающих в этой системе: алгоритм перевода родного слова на иностранный язык и алгоритм перевода иностранного слова на родной язык. Отметим, что оба алгоритма будут алгоритмами с оракулом.

Алгоритм перевода на иностранный язык.

Шаг 1. Пусть нам задано "родное" слово ω^0. Запускаем алгоритм определения семантического поля, который выдает поле, содержащее слово ω^0 и отражающее нужный нам смысл. Обозначим слова, полученные на выходе через $\omega^0_1, \ldots, \omega^0_k$.

Шаг 2. Находим все слова $\omega_*^1 \in \Omega_1$ отвечающие равенству

$$(\omega_1^0 * \ldots \omega_k^0) * \omega_*^1 =_{\gamma > \delta} \omega_*^1.$$

Согласно со случаем 2.1, описанным выше, γ будет высчитываться по формуле

$$\gamma = f^k\big(f(\alpha_1, T(\omega_1^0, \omega_*^1)), \ldots, f(\alpha_k, T(\omega_k^0, \omega_*^1))\big),$$

где $\omega_1^0 * \ldots * \omega_k^0 =_{\alpha_i} \omega_i^0$, $i = 1, \ldots, k$.

Если на выходе мы будем получать пустое множество, то нечеткое отношение T задано некорректно, так как свойство (4) не выполняется.

Перед тем, как описывать второй алгоритм, определим следующее понятие. Допустим нам задано некоторое иностранное слово ω^1 и подмножество $\Omega_0^* \subset \Omega_0$ множества родных слов. Будем говорить, что слово $\omega^0 \in \Omega_0^*$ является ближайшим переводом слова ω^1 в множестве Ω_0^*, если выполняется условие $T(\omega^0, \omega^1) = \max\{T(\omega_i^0, \omega^1) | \omega_i^0 \in \Omega_0^*\}$.

Алгоритм перевода на родной язык.

Шаг 1. Пусть нам задано иностранное слово ω^1, для которого нужно найти перевод. Вычисляем $\omega^1 * \omega^1$. Обозначим через $X = \overline{\omega^1 * \omega^1} = \{\omega_1^0, \ldots, \omega_k^0\}$. Дальнейшая задача алгоритма среди слов множества X выделить то семантическое поле, которое выберет оракул.

Шаг 2. Среди слов множества X выберем ближайший перевод слова ω^1. Допустим, что это слово ω_i^0. Занесем его в множество Y и вычислим произведение $\omega_i^0 * \omega_i^0$. Из множества X вычеркнем все слова, принадлежащие множестве $\overline{\omega_i^0 * \omega_i^0}$. Далее, из слов, оставшихся в множестве X, выбираем ближайшее слово к слову ω^1 и повторяем процедуру до тех пор, пока множество X не окажется пустым.

Шаг 3. Обращаемся к оракулу с предложением выбрать одно слово из множества Y. Обозначим его через ω.

Если же множество Y состоит из одного элемента, то необходимости обращаться к оракулу нет. Мы автоматически выбираем это единственное слово.

Далее переходим к алгоритму определения семантического поля с одной только оговоркой, что все вычисляемые там произведения будем ограничивать на множество $\overline{\omega^1 * \omega^1}$. Так, например, вместо множества $\overline{\omega * \omega}$ будем рассматривать множество $\overline{\omega * \omega} \cap \overline{\omega^1 * \omega^1}$. Результат этого алгоритма и будет результатом алгоритма перевода на родной язык.

В алгебраической системе $< \Omega, * >$ возможно задать массу различных алгоритмов, выполняющих те или иные функции, как в родном, так и в иностранном языках. Например, мы можем задать алгоритм определения семантического поля для заданного иностранного слова или алгоритм определения, является ли данное иностранное слово омонимом в своем словаре. Причем последний алгоритм будет алгоритмом без оракула.

§ 4. Многоязычный словарь–переводчик

Пусть нам задана некоторая группа $G = \{g_0, g_1, \dots, g_m\}$ порядка $m + 1$, каждый элемент которой имеет порядок равный 2. Для определенности будем считать, что элемент g_0 является нейтральным элементом этой группы. Существование такой группы для некоторого конкретного m обеспечивает возможность конструирования переводчика, содержащего один родной и m иностранных языков.

Допустим, что нам даны $m+1$ множеств $\Omega_i = \{\omega_1^{g_i}, \dots, \omega_{N_i}^{g_i}\}$, $i = 0, \dots, m$. Каждое множество Ω_i состоит из слов конкретного естественного языка и никакие два множества не определены в одном и том же естественном языке. Язык, которому принадлежат слова множества Ω_0 будем называть родным языком и будем полагать, что на Ω_0 определено отношение S, и, значит, работает алгоритм определения семантического поля. Все остальные языки будем называть иностранными языками. Также будем полагать, что задано m нечетких отношений перевода $T_i \subset \Omega_0 \times \Omega_i$, $i = 1, \dots, m$, аналогичных описанному в предыдущем пункте отношению перевода T.

Пусть $\Omega = \Omega_0 \cup \Omega_1 \cup \dots \cup \Omega_m$. Структура алгебраической системы $< \Omega, * >$ будет задаваться с помощью следующей процедуры. Для любых $\omega_p^{\varepsilon_p}, \omega_{i_1}^{\varepsilon_1}, \dots, \omega_{i_k}^{\varepsilon_k} \in \Omega$, $\varepsilon_j \in G$, имеем

$$\omega_{i_1}^{\varepsilon_1} * \ldots * \omega_{i_n}^{\varepsilon_n} =_\alpha \omega_p^{\varepsilon_p},$$

где $\alpha = 0 \Leftrightarrow \varepsilon_p \neq \varepsilon_1 + \ldots + \varepsilon_n$.

Нетрудно убедиться, что структура $< \Omega, * >$ является размытой группой со скелетом, изоморфным группе G.

Интерпретация только что описанной операции будет подразделяться на три случая. Допустим, что мы уже совершили $n-1$ действие. Согласно свойствам размытых групп, в результате выполнения этих действий мы получим все элементы некоторого множества Ω_p, $p = 0, \ldots, m$ (каждый элемент со своим приоритетом). Выполнение n-го действия будет осуществлять одну из следующих процедур.

1°. Подбор семантически подобных слов в языке Ω_p. Как и в случае двуязычного переводчика, эта процедура осуществима, если в качестве n-го действия выбрано умножение на элемент из Ω_0. Весовые значения операции будут высчитываться так же, как в случаях 1.1 и 1.2, описанных в предыдущем параграфе.

2°. Перевод слов из языка Ω_p в некоторый другой язык. Если $\Omega_p = \Omega_0$, т.е. после выполнения $n-1$ действия мы оказались в родном языке, то, выполняя умножение на некоторое иностранное слово ω^{g_k}, мы получим перевод заданных родных слов на язык Ω_k. Весовые значения этой операции будем подсчитывать аналогично случаю 2.1 предыдущего параграфа.

Если же $\Omega_p \neq \Omega_0$, то, для обеспечения действия перевода, мы должны умножать на некоторое иностранное слово ω^{g_p}, принадлежащее языку Ω_p. Результатом выполнения этой операции будет перевод заданных иностранных слов на родной язык. Весовые значения этой операции будем подсчитывать аналогично случаю 2.2 предыдущего параграфа.

3°. Некорректное задание входных параметров перевода. Допустим, что в результате выполнения $n-1$ операции мы попали в некоторый иностранный язык Ω_p. Мы будем говорить, что входные параметры заданы некорректно, если в качестве n-го действия будет рассматриваться умножение на некоторое иностранное слово ω^{g_k}, не принадлежащее языку Ω_p. Дадим формальное описание этой процедуры.

Итак, мы имеем $\omega_{i_1}^{\varepsilon_1} * \ldots * \omega_{i_n}^{\varepsilon_n} =_{\alpha_j} \omega_j^{g_p}$, $j = 1, \ldots, N_p$.

Тогда $(\omega_{i_1}^{\varepsilon_1} * \ldots * \omega_{i_n}^{\varepsilon_n}) * \omega^{g_k} =_\delta \omega_j^{g_p + g_k}$, $j = 1, \ldots, N_{p+k}$.

Весовые значения операции задаются таким образом, чтобы каждая структура $< \Omega_0 \cup \Omega_i, * >$, $i = 1, \ldots, n$, являлась структурой двуязычного переводчика. Тогда алгоритмы перевода с родного языка на иностранный язык и с иностранного языка на родной язык будут аналогичны алгоритмам, описанным в предыдущем параграфе.

Опишем алгоритм перевода с одного иностранного языка на другой иностранный язык. Допустим нам задано слово ω^{g_r} принадлежащее языку Ω_r, и ставиться задача перевести его на язык Ω_S.

Шаг 1. Запускаем алгоритм перевода слова ω^{g_k} на родной язык. Результатом этого алгоритма будет семантическое поле слов родного языка, отражающее смысл слова ω^{g_r}. Обозначим слова этого ряда через $\omega_1^{g_0}, \ldots, \omega_k^{g_o}$.

Шаг 2. Находим множество слов вида $\omega_x^{g_s}$, отвечающих равенству $\omega_1^{g_0} * \ldots * \omega_k^{g_0} * \omega_x^{g_S} =_{\alpha > \delta} \omega_x^{g_S}$, которые и будут являться переводом исходного слова на язык Ω_S.

Заметим, что выполнение свойств (4) для отношения T_S обеспечивает на выходе непустое множество переводов.

Замечание 3. Все описанные выше алгоритмы были бы намного проще, если бы мы изначально в каждом иностранном языке определили алгоритм определения семантического поля. А не сделали мы это из тех соображений, что даже в родном языке очень трудно объективно определить структуру семантических полей (т.е. задать отношение S). А познания в иностранном языке у любого эксперта более ограничены, чем познания в родном языке. Наша же система, обладая минимумом входной информации, сама делает некоторые выводы о структуре иностранного языка и его взаимосвязей с родным языком.

Замечание 4. Построение структуры, содержащей $n + 1$ различных языков, гораздо удобнее, чем если бы мы для этих же целей рассматривали множество двуязычных структур. Для того, чтобы обеспечить перевод с одного из $n + 1$ языка на любой другой, мы должны были бы определить $n(n + 1)$ двуязычную структуру. Более того, по крайней мере n языков хотя бы один раз должны были бы выступать в роли родного языка.

После разделения в 1991 году Советского Союза образовался целый ряд государств, обладающих трехязычной структурой: государственный язык, межнациональный язык, международный язык. Это событие делает актуальным создание трехязычного переводчика. Однако, очевидно, что не существует трехэлементной группы, порядок каждого элемента которой равен 2. Эта проблема может быть решена следующим образом.

Ни у кого не вызывает возражений существование четырехэлементной группы, порядок каждого элемента которой равен 2. Это делает возможным построение четырехязычного переводчика. Только в качестве третьего иностранного языка мы возьмем так называемый "вырожденный" язык, состоящий всего из одного слова ω^*. Нечеткое отношение T_3 зададим следующим образом $\forall \omega^{g_0} \in \Omega_0(T(\omega^{g_0}, \omega^*) = \delta)$, т.е. мы нарушаем выполнение свойства (4). Таким образом, при попытке произвести какие–либо переводы, связанные с этим языком, мы будем получать на выходе пустое множество. На работе же остальных алгоритмов этот факт ни как не будет сказываться и система будет работать бесперебойно.

Обращаясь к классической теории групп [9], несложно доказать, что для любого заданного натурального числа m существует группа конечного порядка, большего или равного m, каждый элемент которой имеет порядок равный двум. Таким образом, вводя нужное число "вырожденных" языков, мы можем построить переводчик для любого числа языков.

Л и т е р а т у р а

1. ZADEH L.A. Fuzzy sets //Inform. and control. — 1965. — № 8. — P. 338–353.

2. АРАПОВ М.В. Связь между полисемией и употребительностью слова //Лингвистические задачи и обработка данных на ЭВМ. — М., 1987. — С. 171–192.

3. АХМАНОВА О.С. Словарь лингвистических терминов. — М.: Сов.энциклопедия, 1996. — 608 с.

4. ВЕЖБИЦКАЯ А. Язык. Культура. Познание. — М.: Русские словари, 1996. — 416 с.

5. АЛЕКСАНДРОВА З.Е. Словарь синонимов русского языка. — М.: Русский язык, 1986. — 600 с.

6. DOMBI J. A general class of fuzzy operators, the De Morgan class of fuzzy operators and fuzziness measures induced by fuzzy operators //Fuzzy Sets and Sistems. — 1982. — Vol. 8. — P. 55–73.

7. ДОБРИЦА В.П., ЯХЪЯЕВА Г.Э. О группах с нечеткими операциями //Структурные и сложностные проблемы вычислимости. — Новосибирск, 1999 — Вып. 165: Вычислительные системы. — С. 127–138.

8. ЯХЪЯЕВА Г.Э. Изоморфизмы размытых групп. //Вестник АГУ: сер. физико-математических наук. — Алматы. — 2000. — № 1. — С. 90–98.

9. КАРГАПОЛОВ М.И., МЕРЗЛЯКОВ Ю.И. Основы теории групп. — М.: Наука. — 1982. — 288 с.

Поступила в редакцию
22 ноября 2001 года