

## РОЗДІЛ ІХ. ПРИКЛАДНА ЛІНГВІСТИКА: НАПРЯМИ Й АСПЕКТИ ДОСЛІДЖЕННЯ

Дмитрий Воевудский, Алексей Кретов, Геннадий Селезнев

ББК 81.2  
УДК 811.111**КВАНТИТАТИВНЫЕ ПОКАЗАТЕЛИ МНОГОЗНАЧНОСТИ  
В НИДЕРЛАНДСКОМ, НЕМЕЦКОМ И АНГЛИЙСКОМ СЛОВАРЯХ**

У статті встановлено й зіставлено епідигматичний потенціал лексичних одиниць нідерландської, німецької й англійської мов, що ґрунтується на даних про багатозначність.

Ключові слова: епідигматика, епідигматична стратифікація, коефіцієнт багатозначності, нідерландська мова.

Одним из актуальных направлений современной системной лингвистики является параметрический анализ лексики, разработанный В.Т. Титовым [Титов, 2004: 6-17]. Одним из этапов анализа является эпидигматическая стратификация лексики.

В предлагаемой работе исследуются количественные показатели многозначности на примере трех западногерманских языков – нидерландского, немецкого и английского. Источниками исследования послужили так называемые «малые» двуязычные словари – Дренясова Т. Н., Миронов С. А. Карманный нидерландско-русский словарь (около 7000 слов); Ахманова О.С. Англо-русский словарь (краткий) (около 8000 слов); Липшиц О. Д. Краткий немецко-русский словарь (около 9000 слов).

Стандартный учет эпидигматического потенциала нидерландских слов, основанный на данных о многозначности, позволил получить следующие результаты:

Количество значений	Количество слов	Накопленное количество слов	Вес ранга
9	1	1	0,99986
8	1	2	0,99971
7	1	3	0,99957
6	6	9	0,99871
5	39	48	0,99309
4	104	152	0,97813
3	381	533	0,92331
2	1261	1794	0,74187
1	5156	6950	0,00000

В наглядном виде данная информация представлена на графике:

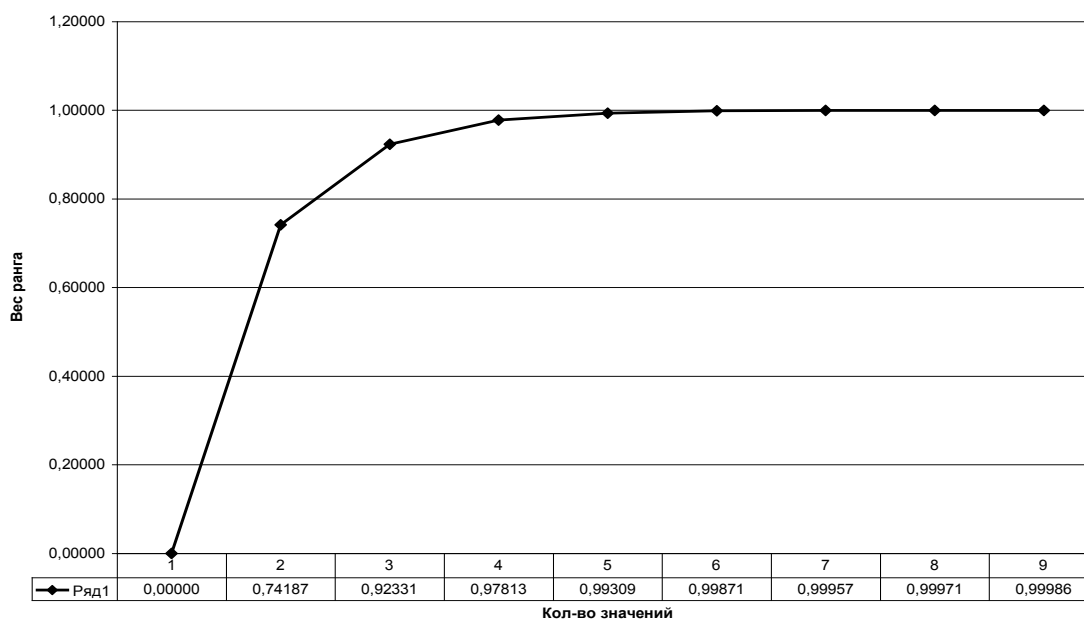


Рис. 1. Динамика Ф-веса в зависимости от числа значений в нидерландско-русском словаре.

Что касается английского языка, то его эпидигматический потенциал может быть представлен следующим образом:

Количество значений	Количество слов	Накопленное количество слов	Вес ранга
7	1	1	0,99988
6	3	4	0,99951
5	19	23	0,99719
4	73	96	0,98828
3	358	454	0,94458
2	1536	1990	0,75708
1	6203	8192	0,00000

В наглядном виде данная информация представлена на графике:

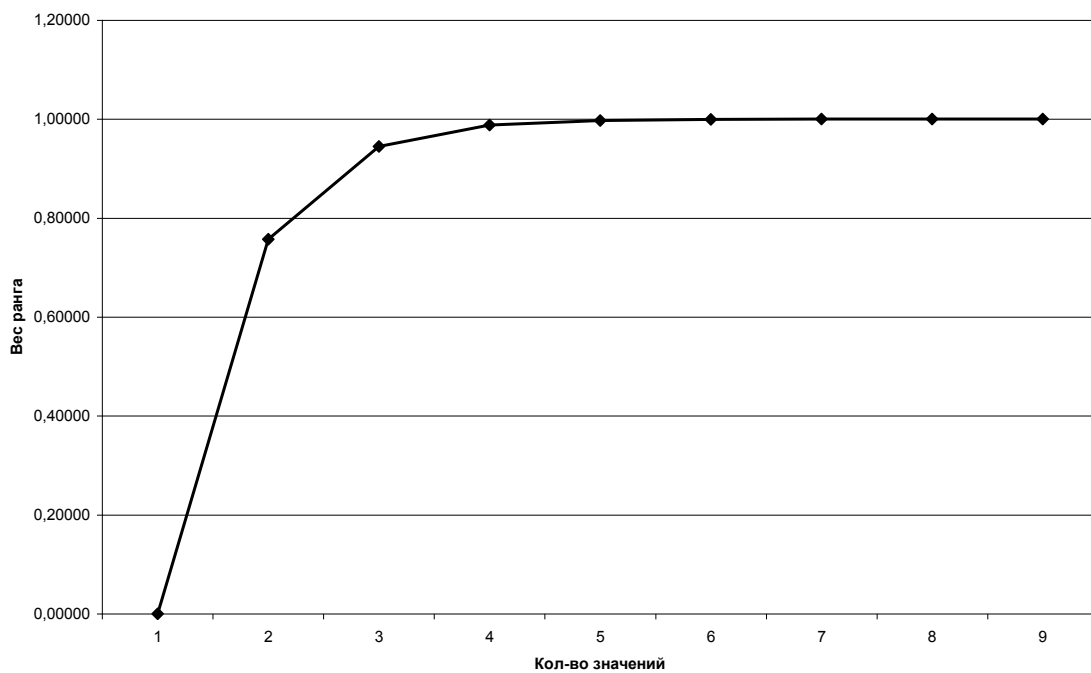


Рис. 2. Динамика Ф-веса в зависимости от числа значений в англо-русском словаре.

Эпидигматический потенциал немецкого языка может быть представлен следующим образом [Казакова 2008: 85]:

Количество значений	Количество слов	Накопленное количество слов	Вес ранга
8	2	2	0,99977
7	1	3	0,99966
6	1	4	0,99954
5	18	22	0,99749
4	62	84	0,99041
3	241	325	0,96289
2	1135	1460	0,83328
1	7297	8757	0,00000

В наглядном виде данная информация представлена на графике:

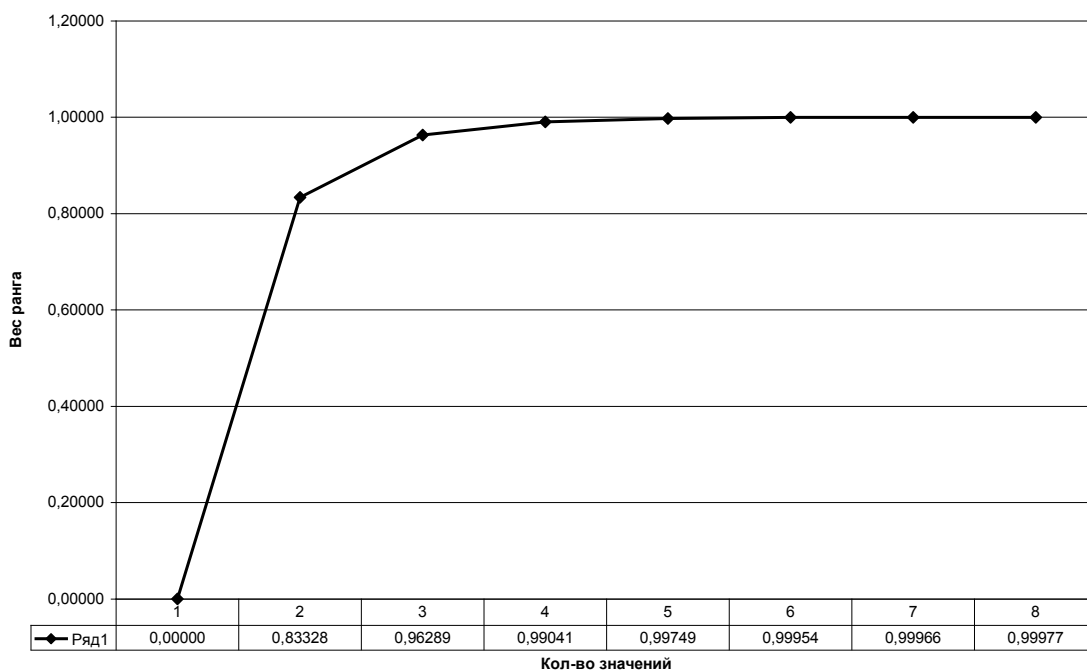


Рис. 3. Динамика Ф-веса в зависимости от числа значений в немецко-русском словаре.

Сравним графики английского, немецкого и нидерландского языков:

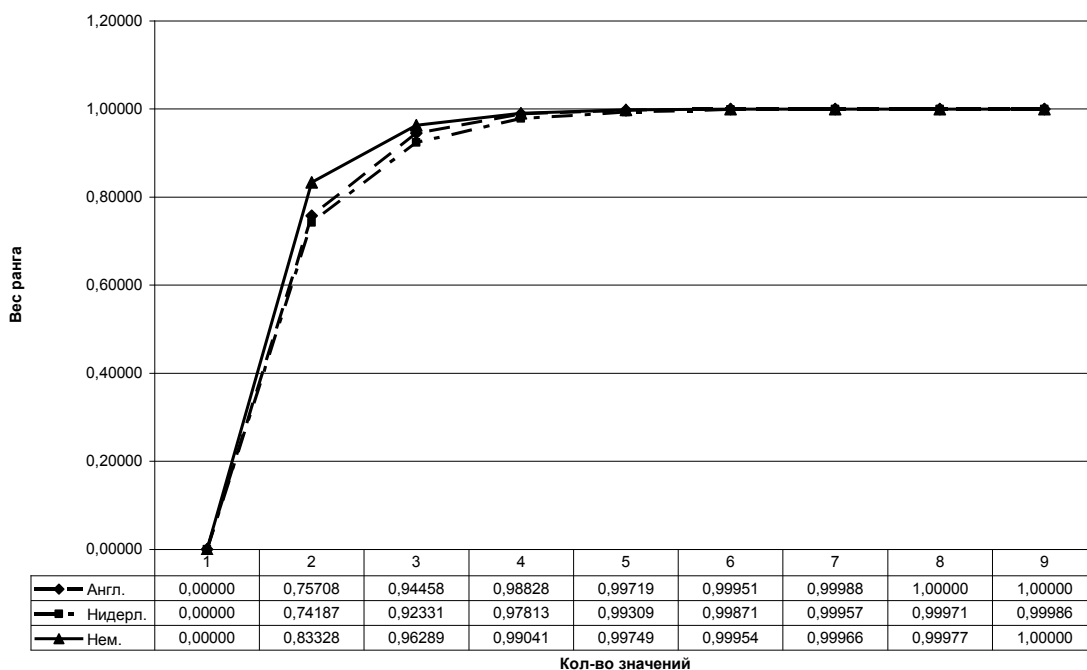


Рис. 4. Динамика Ф-веса в зависимости от числа значений в словарях западногерманских языков.

Теперь мы имеем все необходимое для вычисления коэффициента многозначности. Он представляет собой среднее количество значений, приходящихся на слово. Получается он делением суммарного количества значений на суммарное количество слов в словаре [Титов, 2002: 44]. Значения представлены на гистограмме:

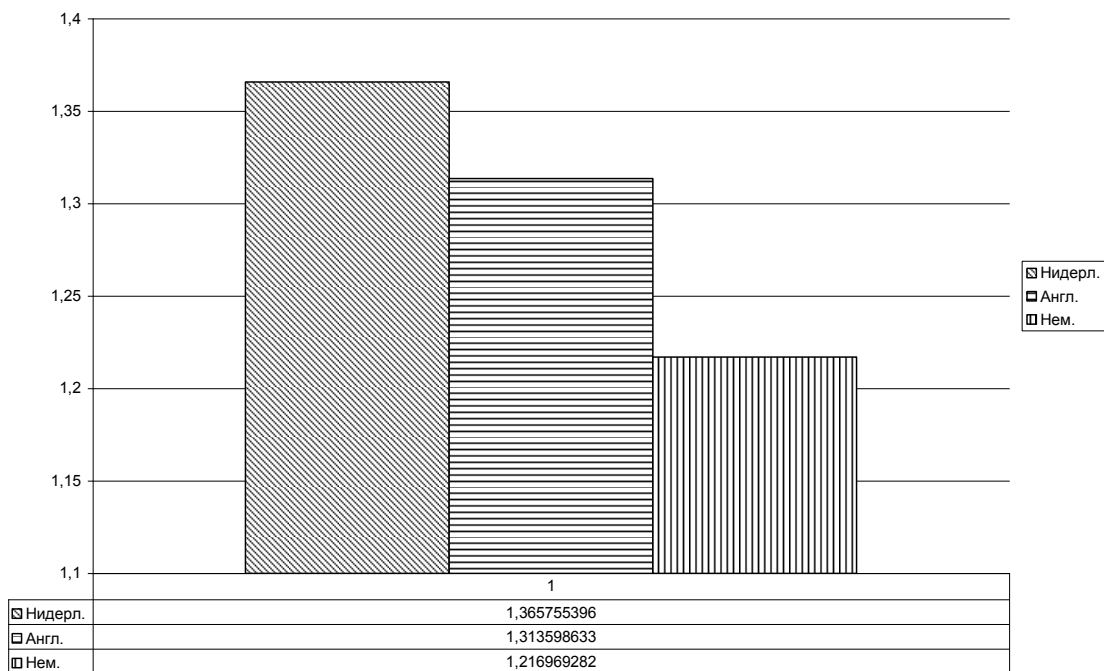


Рис. 5. Значения коэффициента многозначности в словарях западногерманских языков.

Таким образом, коэффициент многозначности нидерландского словаря равен 1,36576, немецкого – 1,21697, а английского – 1,31360, а Ф-вес приведен в нижеследующей таблице.

Кол-во значений слов	Ф-вес ранга в нидерландском	Ф-вес ранга в английском	Ф-вес ранга в немецком
9	0,99986	-	-
8	0,99971	-	0,99977
7	0,99957	0,99988	0,99966
6	0,99871	0,99951	0,99954
5	0,99309	0,99719	0,99749
4	0,97813	0,98828	0,99041
3	0,92331	0,94458	0,96289
2	0,74187	0,75708	0,83328
1	0,00000	0,00000	0,00000

Для увеличения достоверности аппроксимации для некоторых языков не принимались в расчет слова с самым большим числом значений или незначительно изменялось их количество (не более 3 слов для немецкого словаря, т.е. 0.034 % от объема словаря).

Значени й	Нидерландский		Английский		Немецкий	
	Данные	Корректировка	Данные	Корректировка	Данные	Корректировка
1	5156	5156	6203	6203	7297	7297
2	1261	1261	1536	1536	1135	1135
3	381	381	358	358	241	241
4	104	104	73	73	62	62
5	39	39	19	19	18	18
6	6	6	3	3	1	3
7	1	1	1	1	1	
8	1				2	
9	1					

Средствами EXCEL проводилась аппроксимация этих распределений. Наилучший результат достигается при аппроксимации данных экспоненциальным распределением вида

$$N_i = K \exp(-\alpha i), \text{ где}$$

$N_i$  – количество слов в словаре с числом значений равным  $i$ ,  $i = 1, 2, \dots, n$ ;

$\alpha, K$  – подбираемые при аппроксимации показатель экспоненты и масштабный коэффициент.

Язык	K	$\alpha$	Достоверность аппроксимации $R^2$ (%)
Нидерландский	2315 2	1.379 2	0.9913
Английский	2755 6	1.462 3	0.9994
Немецкий	2687 7	1.507 8	0.9960

На представленных диаграммах (Рис. 6-8) численные значения  $N_i$  представлены в логарифмической шкале.

**Распределение слов по числу значений в словаре  
нидерландского языка**

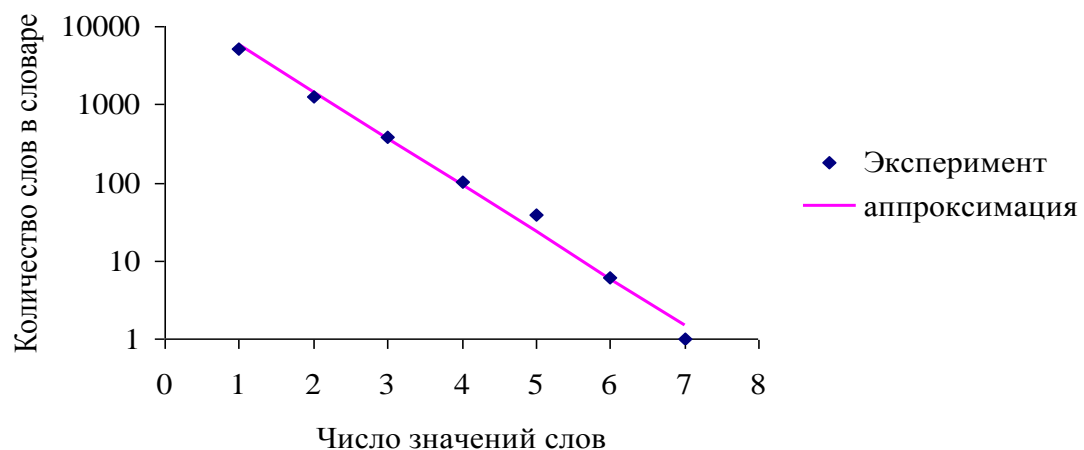


Рис. 6. Распределение слов по числу значений в нидерландско-русском словаре.

**Распределение слов по числу значений в словаре  
английского языка**

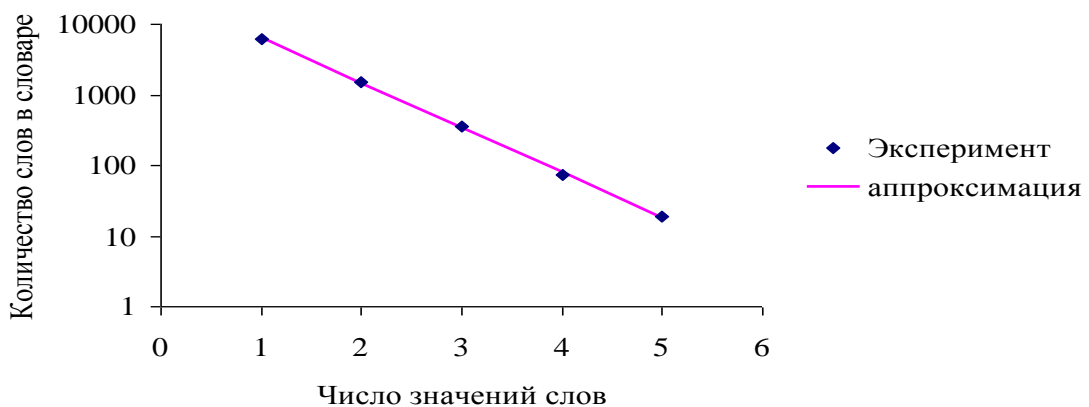


Рис. 7. Распределение слов по числу значений в английско-русском словаре.

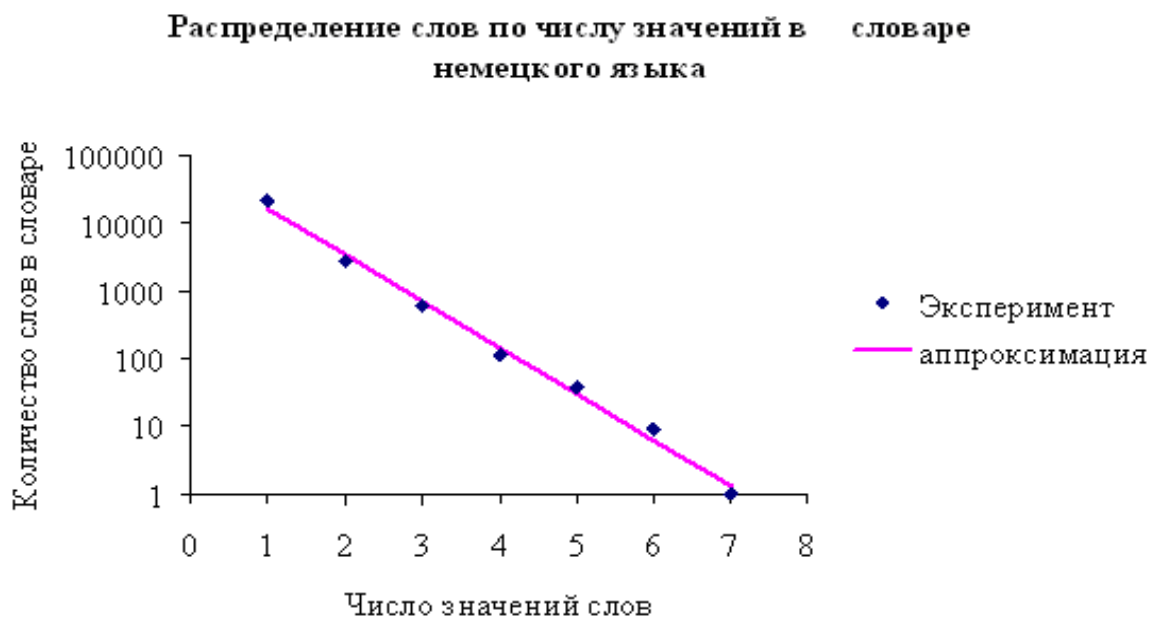


Рис. 8. Распределение слов по числу значений в немецко-русском словаре.

Поскольку коэффициент  $\alpha$  не зависит от объёма словаря, а зависит от языка, он обладает типологической, а следовательно, и классифицирующей значимостью.

В этой связи интересно отметить зеркальную симметричность двух типологических показателей: коэффициента многозначности и коэффициента  $\alpha$  (см. Рис.9).

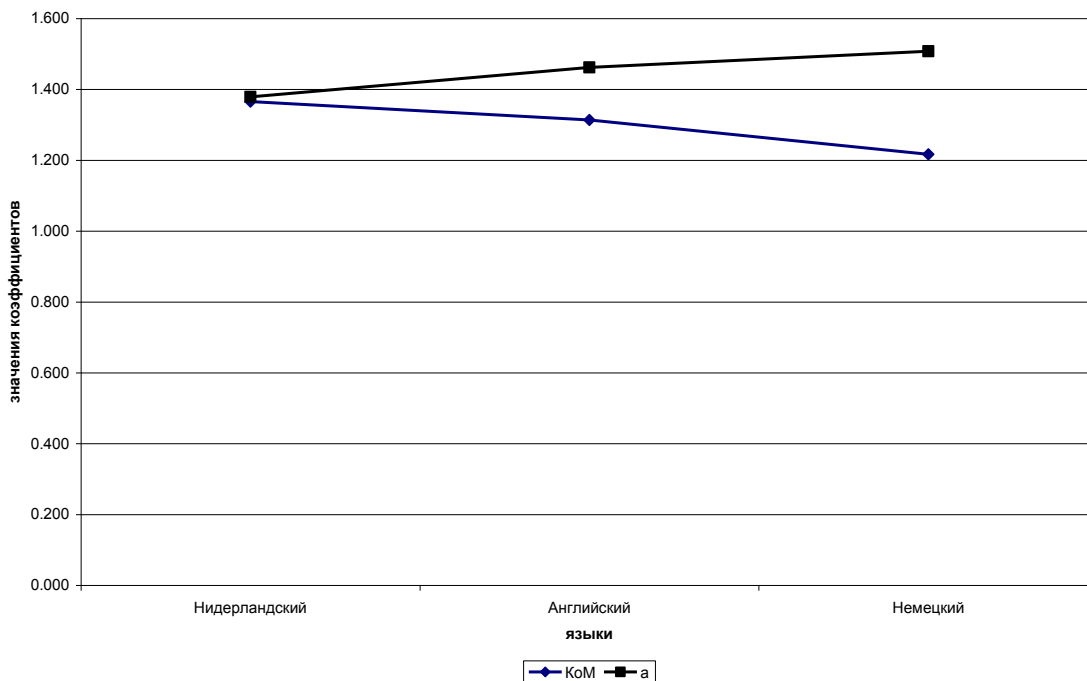


Рис. 9. Значения коэффициентов многозначности и  $\alpha$  в западногерманских языках.

Как показывает Рис. 9, коэффициенты многозначности и  $\alpha$  находятся в отношении зеркальной симметрии: минимальному значению коэффициента многозначности в английском словаре соответствует максимальное значение коэффициента  $\alpha$  и наоборот.

Взаимодействие этих типологически значимых для характеристики лексико-семантических систем параметров представляется нетривиальным результатом предпринятого исследования.

Література

- Ахманова 1993: Ахманова, О. С. Англо-русский и русско-английский словарь (краткий) [Текст] / О.С. Ахманова, Е.А. Уилсон (ред.). – М.: Рус.яз., 1993. – 663 с. – 4000.00 пр. – ISBN 5-200-02044-1.
- Дренясова, Миронов 1977: Дренясова, Т. Н., Миронов, С. А. Карманный нидерландско-русский словарь. Около 7000 слов [Текст] / Т.Н. Дренясова, С.А. Миронов. – М.: Издательство “Русский язык”, 1977. – 392 с. – ISBN 5-200-02780-2: 48.00.
- Казакова 2008: Казакова, Т. А. Параметрический анализ немецкой лексики [Текст]: диссертация. канд. филологических наук: 10.02.04 / Воронеж. гос. ун-т. – Воронеж, 2008. – 187, 66 с.: ил. + автореферат (19 с.). – Библиогр.: с. 171-187. – На правах рукописи.
- Липшиц, Лоховиц 1986: Липшиц, О. Д., Лоховиц, А. Б. Краткий немецко-русский и русско-немецкий словарь [Текст] / О.Д. Липшиц, А. Б. Лоховиц. – М.: Рус. яз., 1986. – 616 с.
- Титов 2002: Титов, В.Т. Общая квантитативная лексикология романских языков [Текст] / В.Т. Титов. – Воронеж: Изд-во Воронеж. гос. ун-та, 2002. – 237 с. – ISBN 5-7455-1227-х.
- Титов 2004: Титов, В.Т. Частная квантитативная лексикология романских языков [Текст] / В.Т. Титов. – Воронеж: Изд-во Воронеж. гос. ун-та, 2004. – 549 с. – ISBN 5-9273-0739-6.

*В данной работе проводится учет и сравнение эпидигматического потенциала лексических единиц нидерландского, немецкого и английского языков, основанный на данных о многозначности.*

*Ключевые слова: эпидигматика, эпидигматическая стратификация, коэффициент многозначности, нидерландский язык.*

*In this article the authors investigate and compare epidigmatic potential of lexemes in Dutch, German and English, based on data of their polysemy.*

*Keywords: epidigmatics, epidigmatic stratification, index of polysemy, Dutch.*

Надійшла до редакції 18 січня 2010 року.

ББК Ш111с51

УДК 81'322

І.Г.Данилюк

**ЛІНГВІСТИЧНЕ ЗАБЕЗПЕЧЕННЯ КОМП'ЮТЕРНИХ СИСТЕМ:  
РЕГУЛЯРНІ ВИРАЗИ В MS WORD 2010**

*У статті розглядаються прикладні аспекти використання регулярних виразів у поширеному текстовому редакторі компанії Microsoft. Наведено галузі застосування РВ у прикладній лінгвістиці, окреслено специфіку РВ у MS WORD 2010, наведено приклади автоматизації певних завдань та процедур роботи з текстом, а також описано можливість проведення окремих лінгвістичних досліджень (насамперед, квантитативних) філологами без глибокої технічної підготовки.*

*Ключові слова: регулярні вирази, текстовий процесор, пошук, заміна, спеціальний символ.*

Використання спеціальних шаблонів для автоматичного чи автоматизованого опрацювання текстових даних є невід'ємним інструментом сучасних текстових процесорів та повсякденною практикою фахівців з розробки прикладного програмного забезпечення, лінгвістів, що опрацьовують мовні дані, численних користувачів мовнозорієнтованого ПЗ. Регулярні вирази (англ. regular expressions, скорочено Regexp, Regex) – це формальна мова пошуку й здійснення маніпуляцій з підрядками в тексті, що ґрунтується на використанні метасимволів (символів-джокерів, англ. wildcard characters) [Фридл 2001: 19]. За іншим визначенням, це форма стандартного запису текстових даних, що використовується в автоматах зі скінченною кількістю станів [Jurafsky 2009: 17].

По суті, РВ – це рядок-зразок (англ. pattern, також його називають «шаблоном», «маскою»), який складається із символів і метасимволів, що й задає правило пошуку.

Джерела регулярних виразів лежать у теорії автоматів і теорії формальних мов. Ці галузі вивчають обчислювальні моделі (автомати) і способи опису й класифікації формальних мов. У 1940-х рр. Уоррен Маккалок і Уолтер Піттс описали нервову систему, використовуючи простий автомат як модель нейрона. Математик Стівен Кліні пізніше описав ці моделі, використовуючи свою систему математичних позначень, яку назвав «регулярними множинами». Кен Томпсон вмонтував їх у редактор QED, а потім у редактор ed під UNIX. У складі цих та інших утиліт (редактор sed і фільтр grep), що постачалися у дистрибутивах UNIX, регулярні вирази зробили прорив в електронній обробці текстів наприкінці XX століття. Багато сучасних мов програмування мають вбудовану підтримку регулярних виразів: серед них Perl, Java, PHP, Javascript, мови платформи .NET Framework, Python та ін. [Смит 2006: 12-34].

Регулярні вирази використовуються численними текстовими процесорами, редакторами й утилітами для пошуку й підстановки тексту.