

## ОПЫТ СОЗДАНИЯ ЧАСТОТНОГО СЛОВАРЯ СОВРЕМЕННОЙ АНГЛИЙСКОЙ ЭКОНОМИЧЕСКОЙ ЛЕКСИКИ

Включение России в глобальное экономическое пространство, а также существенная структурная перестройка в мировой экономике ставят новые задачи исследовательского и практического характера перед сообществом профессионалов, которые заняты подготовкой специалистов, способных эффективно работать в области экономики. Это не только текущие проблемы международного экономического взаимодействия, но и задачи быстрой переработки и усвоения стремительно устаревающей информации экономического характера. Одновременно ощущается возрастание значимости и востребованности экономических знаний. По нашим оценкам, в 90-е годы прошлого века доля экономической литературы составляла в среднем от 10 до 30% от всего потока научных публикаций. Растет и цена правильного понимания термина. Так, например, в 2000 г. рынок сбережений и инвестиций Великобритании недополучил 10 млрд фунтов стерлингов из-за того, что 67 % людей не знали, что такое ISA (ISA — Individual Savings Account — Индивидуальный сберегательный счет).

В настоящее время мировая экономическая наука представляет собой достаточно сложное образование, о чем, в частности, свидетельствует классификация публикаций, созданная Американской экономической ассоциацией и применяемая всеми ведущими экономическими журналами. В этой классификации 19 макроразделов, 125 мезоразделов и 724 микрораздела. Исследования показывают, что одни и те же слова используются с разной частотой и смысловым наполнением в разных подразделах. В силу упоминавшегося выше динамизма хозяйственной жизни происходит быстрое обновление экономической лексики. Появление электронных словарей типа Lingvo существенно облегчает работу по переводу терминологической лексики. Однако, с точки зрения специалистов-экономистов, которым приходится постоянно работать с современной англоязычной литературой, существующие англо-русские специализированные словари нуждаются в постоянном совершенствовании и обновлении. По нашему мнению, одним из путей решения этой проблемы на новом качественном уровне может стать *разработка частотно-ассоциативных словарей экономической лексики*. Такие словари могут не только стать эффективным помощником для более быстрого освоения конкретной предметной области, дополняя средства типа Lingvo, Promt и др., но и послужить источником надежной информации для исследовательской и практической работы, связанной с подготовкой учебных и лексикографических материалов в системе обучения переводчиков и экономистов.

Сформированный на базе НГУ творческий коллектив в настоящее время занят концептуальной разработкой макроструктуры будущего словаря, а также апробацией и совершенствованием созданных специально для этой задачи технологий. Полученный в ходе этой работы опыт позволяет предложить следующую схему достижения основной цели.

1. В качестве основы берется крупнейшая в мире библиографическая англоязычная база EconLit-AEA, которая дополняется полнотекстовыми публикациями журнальных статей. При помощи программы, разработанной А.С. Шевцовым, проводится подсчет числа словоформ по каждому из 724 предметных микроразделов классификации Американской экономической ассоциации. Глубина и тщательность про-

работки отдельного раздела определяются публикационной значимостью раздела в период с 1992 по 2004 год.

2. Полученные массивы очищаются от «мусора» и распределяются по рубрикам, которые содержат а) служебные слова; б) иноязычные вкрапления; в) слова общелитературного языка; г) общенаучную лексику; д) географические названия; е) имена собственные; ж) собственно экономические термины; з) значимые сокращения. Рассчитываются частоты словоупотреблений по очищенному массиву в целом и по отдельным выделенным рубрикам как для отдельных слов, так и нарастающим итогом по частоте словоупотребления.

3. Экономические термины, которые критически значимы для адекватного понимания текстов по соответствующим предметным областям, получают расширенное толкование с учетом полисемии (возможен расчет частотности каждого значения), моделируются семантические имплицативные сети, отражаемые в специальных ассоциативных таблицах, даются гиперссылки на наиболее авторитетные и значимые материалы в электронных библиотеках. В этой работе используются достижения современной психолингвистики и когнитивной психологии, связанные с распознаванием и обработкой информации, которая представлена средствами естественного языка.

Так, например, в период с 1992 по 2004 год «чемпионом» по общему числу публикаций, учтенных в EconLit-AEA, был микрораздел O15 «Человеческие ресурсы, распределение дохода, миграция». В названиях и рефератах статей этого раздела встретилось 30123 словоформы, которые были употреблены 893405 раз. Вместе с тем, как свидетельствуют данные табл. 1, 50% общего числа отмеченных словоупотреблений пришлось всего на 123 слова.

Как и следовало ожидать, первые 15 слов, на долю которых приходится 30,4% словоупотреблений, относятся к служебным словам. А всего к этой группе можно отнести 48 слов, которые специально не выделены в таблице. Из географических наименований наиболее часто встретились Africa, India и China. К этой семантической группе тяготеют также слова countries и south. Курсивом выделены шесть общенаучных слов. И, наконец, семантика оставшихся 64 словоформ в той или иной степени наполнена экономическим содержанием. Эти словоформы и послужат материалом для выделения ассоциатов в имплицативной сетке.

Важно подчеркнуть, что в экономической литературе идут дискуссии относительно содержательной трактовки некоторых из приведенных здесь терминов. Это в частности относится к development, но в еще большей степени к labour, суммарная частота которого в американском (labor) и британском (labour) орфографических вариантах составляет 0,5%, что несколько превышает частотный показатель термина development.

Из наблюдения над таблицей видно, что материалы частотных словарей могут дать основу для построения интересных обучающих лингвистических игр по словообразованию и пониманию смысла терминов. Если, например, добавить к частоте development частоту родственной словоформы developing, то получим 0,646, что выведет "развитие" на первое место и поставит впереди слова are. Но are может опять выйти вперед, если "подтянуть" к нему все формы глагола to be. Но и здесь возможен выбор между приглагольной частицей и предлогом в общем объеме частотности (2%) to.

Таблица 1. Первые 123 слова из микрораздела O15 «Человеческие ресурсы, распределение дохода, миграция» на долю которых приходится 50% общего числа словоупотреблений (ранжирование по мере уменьшения частоты словоупотреблений)

Слово	ЧС*	P*	PS*	Слово	ЧС	P	PS	Слово	ЧС	P	PS
the	63764	7,137	7,137	<b>labour</b>	2003	0,224	38,913	<i>article</i>	1196	0,134	45,663
of	42915	4,804	11,941	<b>employment</b>	2003	0,224	39,137	<b>India</b>	<b>1186</b>	0,133	45,795
and	38890	4,353	16,294	we	2001	0,224	39,361	<b>south</b>	<b>1170</b>	0,131	45,926
in	32517	3,640	19,933	<b>human</b>	1977	0,221	39,582	was	1162	0,130	46,056
to	18612	2,083	22,017	more	1899	0,213	39,795	<b>distribution</b>	1146	0,128	46,185
a	14413	1,613	23,630	not	1869	0,209	40,004	<b>case</b>	1127	0,126	46,311
for	8940	1,001	24,631	these	1796	0,201	40,205	how	1115	0,125	46,436
is	8581	0,960	25,591	<i>study</i>	1762	0,197	40,402	<b>role</b>	1088	0,122	46,557
that	8562	0,958	26,549	<b>market</b>	1745	0,195	40,598	both	1083	0,121	46,679
on	8106	0,907	27,457	<b>capital</b>	1705	0,191	40,788	<b>households</b>	1071	0,120	46,798
this	6435	0,720	28,177	<b>analysis</b>	1694	0,190	40,978	other	1068	0,120	46,918
are	5457	0,611	28,788	or	1633	0,183	41,161	two	1037	0,116	47,034
with	5252	0,588	29,376	<b>household</b>	1621	0,181	41,342	<b>child</b>	1032	0,116	47,150
from	4584	0,513	29,889	<b>impact</b>	1591	0,178	41,520	<b>changes</b>	1025	0,115	47,264
as	4390	0,491	30,380	<b>sector</b>	1552	0,174	41,694	among	1018	0,114	47,378
<b>development</b>	4362	0,488	30,868	<b>poor</b>	1495	0,167	41,861	<b>international</b>	1012	0,113	47,492
<b>poverty</b>	4326	0,484	31,353	<b>Africa</b>	<b>1490</b>	0,167	42,028	<b>political</b>	1011	0,113	47,605
by	4322	0,484	31,836	<b>gender</b>	1432	0,160	42,188	<b>work</b>	1009	0,113	47,718
<b>economic</b>	3924	0,439	32,276	<b>model</b>	1429	0,160	42,348	<b>level</b>	1005	0,112	47,830
an	3788	0,424	32,700	<b>developing</b>	1408	0,158	42,506	<b>women's</b>	968	0,108	47,939
<b>social</b>	3366	0,377	33,076	than	1401	0,157	42,663	<b>fertility</b>	965	0,108	48,047
<i>paper</i>	3148	0,352	33,429	<b>new</b>	1389	0,155	42,818	<i>examines</i>	960	0,107	48,154
<b>income</b>	3114	0,349	33,777	<i>results</i>	1383	0,155	42,973	such	956	0,107	48,261
<b>growth</b>	3001	0,336	34,113	also	1381	0,155	43,128	there	950	0,106	48,367
it	2823	0,316	34,429	been	1351	0,151	43,279	<b>use</b>	941	0,105	48,473
<b>countries</b>	<b>2777</b>	0,311	34,740	<b>population</b>	1345	0,151	43,429	<b>trade</b>	927	0,104	48,576
be	2726	0,305	35,045	<b>urban</b>	1324	0,148	43,578	<b>family</b>	920	0,103	48,679
<b>education</b>	2686	0,301	35,346	<b>evidence</b>	1282	0,143	43,721	most	906	0,101	48,781
have	2650	0,297	35,642	<b>migration</b>	1282	0,143	43,865	<b>state</b>	904	0,101	48,882
<b>health</b>	2500	0,280	35,922	some	1279	0,143	44,008	<b>survey</b>	901	0,101	48,983
has	2466	0,276	36,198	<b>public</b>	1269	0,142	44,150	over	901	0,101	49,084
<b>labor</b>	2462	0,276	36,474	<b>effects</b>	1259	0,141	44,291	were	893	0,100	49,184
<b>rural</b>	2429	0,272	36,746	<b>using</b>	1245	0,139	44,430	<b>effect</b>	884	0,099	49,283
<b>between</b>	2399	0,269	37,014	<b>wage</b>	1245	0,139	44,569	<b>index</b>	879	0,098	49,381
at	2247	0,252	37,266	its	1242	0,139	44,709	<b>different</b>	873	0,098	49,479
which	2217	0,248	37,514	<b>policies</b>	1235	0,138	44,847	<b>change</b>	871	0,097	49,576
their	2207	0,247	37,761	but	1233	0,138	44,985	<b>important</b>	868	0,097	49,673
<b>policy</b>	2129	0,238	37,999	<b>world</b>	1232	0,138	45,123	may	866	0,097	49,770
<i>data</i>	2105	0,236	38,235	can	1218	0,136	45,259	<b>government</b>	865	0,097	49,867
<b>women</b>	2032	0,227	38,462	<b>workers</b>	1210	0,135	45,394	they	864	0,097	49,964
<b>inequality</b>	2022	0,226	38,689	<b>economy</b>	1200	0,134	45,529	<b>China</b>	<b>858</b>	0,096	50,060

ЧС — число словоупотреблений; P — удельный вес данной словоформы в общем числе словоупотреблений, в %; PS — удельный вес числа словоупотреблений нарастающим итогом, в %.

Анализ словоупотреблений в микроразделе G12 "Оценка активов" также показывает, что артикли и некоторые другие служебные слова прочно занимают первые

места по частотности, но наиболее часто встречающиеся содержательные термины несколько отличаются по своему набору от состава соответствующей рубрики в других подразделах. Здесь мы находим: stock, market, price, volatility и т.д. (см. табл. 2 ниже).

Таблица 2. Первые 96 слов из микрораздела G12 «Оценка активов», на долю которых приходится 50% общего числа словоупотреблений (ранжирование по мере уменьшения частоты словоупотреблений)

Слово	ЧС*	P*	PS*	Слово	ЧС*	P*	PS*	Слово	ЧС*	P*	PS*
the	67888	7,596	7,596	results	2799	0,313	38,508	at	1604	0,179	45,604
of	38744	4,335	11,931	trading	2722	0,305	38,812	when	1597	0,179	45,783
and	28628	3,203	15,134	models	2523	0,282	39,095	value	1582	0,177	45,960
in	21201	2,372	17,506	not	2510	0,281	39,375	analysis	1562	0,175	46,135
to	17513	1,959	19,465	pricing	2497	0,279	39,655	capital	1466	0,164	46,299
a	16611	1,859	21,324	financial	2406	0,269	39,924	performance	1462	0,164	46,462
that	12343	1,381	22,705	using	2295	0,257	40,181	our	1458	0,163	46,625
is	11033	1,234	23,939	information	2285	0,256	40,436	show	1418	0,159	46,784
for	9687	1,084	25,023	data	2178	0,244	40,680	index	1388	0,155	46,939
stock	8999	1,007	26,030	which	2173	0,243	40,923	period	1388	0,155	47,095
market	8800	0,985	27,014	these	2169	0,243	41,166	changes	1365	0,153	47,247
on	8180	0,915	27,930	equity	2127	0,238	41,404	both	1333	0,149	47,397
are	7824	0,875	28,805	have	2097	0,235	41,638	new	1312	0,147	47,543
this	7527	0,842	29,647	return	2060	0,230	41,869	bond	1305	0,146	47,689
we	6700	0,750	30,397	it	1958	0,219	42,088	two	1296	0,145	47,834
with	6637	0,743	31,139	find	1920	0,215	42,303	investors	1288	0,144	47,978
returns	6574	0,736	31,875	exchange	1905	0,213	42,516	has	1283	0,144	48,122
by	5013	0,561	32,436	stocks	1896	0,212	42,728	effects	1282	0,143	48,265
model	4891	0,547	32,983	rate	1888	0,211	42,939	over	1277	0,143	48,408
an	4424	0,495	33,478	can	1874	0,210	43,149	significant	1243	0,139	48,547
markets	4202	0,470	33,948	also	1818	0,203	43,352	but	1224	0,137	48,684
price	4005	0,448	34,396	than	1787	0,200	43,552	real	1171	0,131	48,815
as	3930	0,440	34,836	or	1752	0,196	43,748	portfolio	1143	0,128	48,943
from	3870	0,433	35,269	empirical	1725	0,193	43,941	some	1141	0,128	49,071
risk	3783	0,423	35,692	time	1707	0,191	44,132	tests	1136	0,127	49,198
prices	3459	0,387	36,079	interest	1705	0,191	44,323	earnings	1103	0,123	49,321
volatility	3406	0,381	36,460	more	1699	0,190	44,513	other	1095	0,123	49,444
paper	3387	0,379	36,839	study	1688	0,189	44,702	structure	1077	0,121	49,564
be	3208	0,359	37,198	rates	1623	0,182	44,884	expected	1067	0,119	49,684
between	3051	0,341	37,540	effect	1615	0,181	45,064	investment	1048	0,117	49,801
asset	2995	0,335	37,875	firms	1613	0,180	45,245	there	1045	0,117	49,918
evidence	2859	0,320	38,195	their	1608	0,180	45,425	they	1019	0,114	50,032

\* См. сноску к табл. 1.

Вероятно, суммарные данные по всем разделам, подразделам и рубрикам позволят сделать лингвостатистические выводы, необходимые для составления толковых и переводных экономических словарей, отражающих современное словоупотребление в данной сфере. А разработанные творческим коллективом программные материалы и технологическая база подготовят почву для постоянного обновления этих данных.