Cambridge ELT Readers Project

Wordlist Database Report

David Eastment
November 29, 1994

Setting up the Wordlist Database: Report

1.   Introduction

The aim of this part of the ELT Readers Project was to take existing wordlists for readers and convert them into database form.

In an earlier report (July 9th 1994), it was suggested that, in order to be able to analyze the lexical content of existing readers, we needed two databases: one containing the wordlists themselves (the subject of this Report) and the other the actual texts (or at least a subset of them).

The July report suggested that the requirement for the database of wordlists was for "a simple flatform database: a complete alphabetical listing" of all the words from publishers' wordlists, Council of Europe wordlists, and possibly the Cambridge English Lexicon.

This database has now been created, without Hindmarsh, but with the inclusion of wordlists from Changes and the New Cambridge English Course.

This report outlines how the database was set up, and what initial analysis has revealed. It is accompanied by a disc with compressed ("ZIPPED") files of the database itself and ASCII versions of the wordlists, together with a complete printout of the raw data.

2.   Setting up the Database

The database was created in three phases, as outlined below:

2.1   Typing in lists

A secretary was recruited to type in wordlists from the following sources:

- OUP: "Oxford Bookworms" (1988);
- Collins -CEL (no date);
- Longman "Structural Readers Handbook" (1987);
- Penguin "Simply Stories" (no date);
- Macmillan "Guide to Rangers" (1974);

and, for purposes of comparison:

- Council of Europe: Waystage and Threshold (1990);
- CUP: "Changes" 1 & 2
- CUP: New Cambridge English Course 1 - 4 (1990 - 1993).

Each list was input as a separate wordprocessed file (using Word 6) in the format "Word", "Grammatical Category", "Level" <carriage return>

---

i.e. with comma separators and a hard return at the end of each line. The input process was not straightforward, however, since the originals did not all conform to this format. Some lists (e.g. Longman) were printed incrementally, with each level including items from preceding other levels, whilst others (e.g. Macmillan) omitted grammatical category.

2.2 Phase Two: List Editing

In Phase Two, I carried out preliminary editing of the files: spelling, checking for consistency (e.g. of abbreviations for grammatical category etc) This turned out to be very time-consuming!

It emerged clearly at this point that although the lists clearly drew heavily on each other, they had been drawn up in very different ways. Some, such as the Macmillan list, confined themselves to single words; others included phrases ("the day before yesterday"). Some placed alternatives on a single line ("Dad\Daddy") whilst others listed them as separate items. Variations in meaning were shown by a variety of different conventions...... The list is a lengthy one.

Although I wanted to preserve the originals as far as possible, it was clear that some editing was necessary. I tried to keep this to a minimum. I then saved the original lists as ASCII files.

2.3 Phase Three: Porting into Paradox

I then wrote several scripts to (a) import the files into separate Paradox databases; and (b) homogenize the names and lengths of the fields, delete blank fields etc. A preliminary analysis revealed the following:

Wordlist	Total Items	Verbs	Nouns
Longman	2592	1242	623
Macmillan	2226	1132	646
Penguin	3124	933	1695
CEL	4734	1263	2182
Oxford	3068	842	1225
Changes	1859	333	977
CEC	3534	676	1514
Waystage	928	159	338
Threshold	952	143	594

I then ported all 9 databases into a single database, adopting the structure suggested in the July report:

Word	Category	Longman	Oxford	Penguin	etc	Notes
cancel	v	3	4	4		
car	n	1	1	1		
cashier	v	3	4	0		
etc						

The readers were entered by level, the CUP coursebooks by book, and Threshold and Waystage by a 0 (i.e. 0 if on the list, blank if not).

The publishers' names were reduced to three-letter abbreviations to that data could be viewed without scrolling. Where an exact match was found (identical word, identical category) the appropriate level was entered in the appropriate column. Where the match was not exact, a new record was inserted (in alphabetical order).

In some cases, this worked fine: "abroad, adv" for example, appears as LSR 3, CEL 4, OUP 4, PEN 3 etc. Unfortunately, however, the large number of "similar but not identical" entries resulted in a large number of "similar but not identical" records being opened.

"Grandfather", for example, appears in three separate entries:

- in "grandfather" itself (LSR 2, CHA 1, and THR)
- in "grandfather \ mother" (PEN 2) and
- in "grandmother \ father \ son" (CEC).

Implicitly, it appears in "grand - (son)" (CEL 2) and in "granddaughter etc) (MAC 6).

Appendix "A" illustrates the problem. The report purports to show which words appear in Waystage but in none of the other readers lists. It comes up with a total of 51. Closer examination, however, reveals that "bathroom" appears in PEN 1 as "bath /-room"; that "holiday(s)" appears as "holiday" in all lists etc.

I spent some time wrestling with this problem. Obviously, the database would be more useful and easier to use if it were edited down to a greater degree of homogeny. Indeed, a thorough edit would probably reduce the records by several thousand. On the other hand, such an edit would require several days work. And, given that in the majority of cases the target words are grouped together alphabetically, the problem is probably not as great as at first it appeared.

I therefore saved the completed database as "CUPRAW" (i.e. C.U.P Raw_Data), and printed out the full list of 8122 items (attached).

2.4   Performance

Given a reasonably powerful PC (mine was a 486 with 8 MB of RAM), the database can be interrogated without significant delays. A single-field sort of the whole database takes less than 10 seconds. Locating a record takes about a second (although the database has not been indexed).

3.   Analysis

3.1   *Using the database with texts:*

The value of the database will only really become apparent when used in conjunction with the reader texts.

As a taster, I looked at the first paragraph of "Desert, Mountain, Sea" (Oxford Bookworms 4). This opens with:

It was a beautiful Australian morning, sunny and clear. Not far from Alice Springs, in the centre of this huge country, a young woman started on a journey that would take her 2800 kilometres across the central desert to the western coast of Australia. With her was her dog, Diggity, and four camels. It was the beginning of a very unusual adventure for the young woman, but it was also the end of many months of preparation.

Analysing this against the database, we find:

Word	In OUP List?	In other lists?
Australian	NO	NO
sunny	NO	CEL, THR, CEC, CHA
huge	Yes, but level 5	
desert	Yes, but level 5	
western	NO	CEL
camel	NO	CEL
beginning	NO	CEL, LSR, MAC, CHA, CEC
unusual	NO	CHA, CEC
preparation	NO	LSR, PEN, MAC

There are just over 70 words in this paragraph, of which about 30 might be considered "content words". Of these, we have no less than 7 words which do not figure in the OUP wordlist at all, and 2 which appear, but at a higher level!

3.2   *Core Word List*

An interesting starting point is to query what words are in ALL of the publishers wordlists, and at what level. This is a simple Paradox query, and takes about 30 seconds to produce an answer. The full listing (1023 items) can be found in Appendix "B".

It is important to note that the listing is not completely reliable, for the reasons given in 2.3: "as", for example, does not appear at all, since it is presented so differently in each list.

What is in the list is less interesting than what is not: since most of the lists seem to have been drawn up from other wordlists, it is not surprising that their is a high level of commonality. Odd things stand out: why is "unfortunately" absent, for example?

But more significant is the absence of technology-related items: CD, fax, modem, telecommunications, portable etc. It is interesting that the secretary who typed in the lists felt that they were "old-fashioned"!

I explored this farther by interrogating the database for items which were in Threshold, but not in any of the publishers' wordlists. (Appendix "C") The query illustrates in part how idiosyncratic the Threshold list is ("thingummyjig", "ugh!", "mutton"), but also reveals a range of contemporary items ("addict", "drugs", "CD", "fax", "hi-jack", "walkman") which are absent from the older lists.

3.3   General

The Paradox format lends itself to simple and effective querying: how are entries with "day" handled, for example (Appendix "D"), or "What words are in Longman but no other lists?" (Appendix "E"). These ad hoc queries can throw up useful nuggets of information: the relative dearth of travel-related items, for example, or the fact that only the Council of Europe lists include ordinals beyond "third".

The real utility of the database, however, will only become apparent when used in conjunction with the texts themselves, as indicated in 3.1.

4.   Conclusions and Recommendations

I suggest the following:

4.1   Spend 2-3 days editing the CUPRAW database. Although the existing database is perfectly usable, it does mean that every query carried out on it demands a certain amount of post-editing. A complete edit of the database would be enormously time-consuming, and difficult to justify. A few days, however, could remove the most serious anomalies and provide a genuinely useful resource.

4.2   Begin the process of analysing readers. The "hi-tech" approach to this would involve converting the existing Paradox database into Advanced Revelation and querying the reader texts against the database from within ARev. One could, however, make a perfectly valid start by analyzing a limited number of readers manually, along the lines indicated in 3.1.

David Eastment
December 1, 1994

Query: What words are in Waystage but not in LSR, CEL, OUP, PEN or MAC?

Word	Category	WAY
am	adv	0
baseball	n	0
bathroom	n	0
camping-site	n	0
cannot	v	0
certainly (not)	adv	0
come along	v	0
come home	v	0
congratulations!	int	0
delayed	adj	0
don't	v	0
d ving licence	n	0
eighth	adj	0
eleventh	adj	0
er.....	int	0
far from	prep	0
fifth	adj	0
fourth	adj	0
full (of)	adj	0
gram(me)	n	0
guest-house	n	0
hallo!	coll	0
holiday(s)	n	0
how old	adj	0
in the centre of	prep	0
kilo(gramme)	n	0
light(s)	n	0
monthly	adj	0
Mrs	n	0
Ms	n	0
m : not	v	0
ninth	adj	0
not always	adv	0
not enough	adj	0
of course (not)	adv	0
pennies	n	0
raincoat	n	0
seventh	adj	0
sincerely	adv	0
sit (down)	v	0
sixth	adj	0
sports	n	0
tenth	adj	0
that (one)	pron	0
this (one)	pron	0
thunderstorm	n	0
twelfth	adj	0
underwent	v	0

A1

Query: What words are in Waystage but not in LSR, CEL, OUP, PEN or MAC?

Word	Category	WAY
weekday	n	0
windy	adj	0
working-hours	n	0

Total: 51

A2

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
a	det	1	1	1	1	1
above	prep	1	2	2	1	2
abroad	adv	3	4	4	3	5
accept	v	3	3	3	4	5
accident	n	2	3	1	1	4
accuse	v	4	5	4	4	6
across	prep	1	1	1	1	1
add	v	3	3	3	3	1
address	n	1	3	2	2	4
admire	v	4	4	6	4	5
adventure	n	4	2	2	2	4
advertise	v	4	4	5	4	6
advice	n	3	5	4	3	4
afford	v	3	4	4	4	4
afternoon	n	1	1	1	1	1
afterwards	adv	3	3	3	3	4
again	adv	1	1	1	1	1
against	prep	2	3	2	2	2
ago	adv	2	2	1	1	3
airport	n	1	3	2	1	5
alive	adj	2	1	1	2	1
all right	coll	1	2	2	2	1
allow	v	4	4	4	4	3
almost	adv	2	3	3	2	2
alone	adj	2	2	2	3	1
along	prep	1	1	1	1	1
already	adv	2	2	2	2	1
although	conj	3	4	3	3	2
always	adv	1	1	1	1	1
among	prep	2	3	3	2	2
amount	n	3	3	6	4	6
and	conj	1	1	1	1	1
anger	n	4	3	5	4	4
angry	adj	1	3	1	1	1
animal	n	1	1	2	1	2
annoy	v	4	5	4	3	6
answer	n	1	1	1	1	1
anxious	adj	4	4	5	5	5
any	det	1	1	2	1	1
anywhere	adv	3	4	2	2	3
appear	v	3	3	3	3	4
apple	n	1	2	1	2	1
argue	v	4	3	3	3	6
arm	n	1	1	1	1	1
army	n	4	2	2	3	4
around	prep	4	3	2	3	2

Total: 46

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
         NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
article	n	4	5	6	5	6
asleep	adj	2	1	2	2	1
at	prep	1	1	1	1	1
at first	adv	2	3	1	2	2
at last	adv	2	3	3	2	1
at once	adv	2	3	1	2	1
attack	v	4	3	5	4	4
audience	n	4	5	5	4	5
aunt	n	2	3	4	2	3
autumn	n	2	2	1	1	3
avoid	v	4	4	4	4	6
awake	adj	2	1	3	2	1
awful	adj	2	5	3	3	3
back	adv	2	1	1	1	1
back	n	1	1	1	1	1
bad	adj	1	1	1	1	1
bag	n	1	1	1	1	3
bake	v	3	4	3	5	4
balance	v	4	4	6	5	6
ball	n	1	1	2	2	2
basket	n	1	4	4	3	3
battle	n	4	3	3	4	5
beach	n	1	4	2	1	3
beard	n	1	4	3	2	2
beautiful	adj	1	1	1	1	1
because	conj	2	1	1	1	1
become	v	3	2	3	3	1
beer	n	1	3	4	2	5
behind	prep	1	1	1	1	2
believe	v	2	2	4	3	3
below	prep	2	2	2	1	1
belt	n	1	4	5	4	6
bend	v	4	3	4	3	5
beside	prep	1	2	2	4	3
best	adj	2	2	1	1	3
better	adj	2	2	1	1	2
better	adv	2	2	1	1	2
between	prep	1	1	1	1	2
beyond	prep	3	3	3	3	3
big	adj	1	1	1	1	1
bird	n	1	1	2	1	1
birth	n	4	3	6	3	6
birthday	n	2	3	2	3	3
bite	v	2	2	3	2	1
black	adj	1	1	1	1	1
blame	v	4	4	4	3	4
blanket	n	2	3	3	3	5

Total: 93

B2

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
         NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
blind	adj	3	4	5	3	3
blood	n	2	2	2	2	2
blow	v	2	3	4	2	2
blue	adj	1	1	1	1	1
boat	n	1	1	2	1	1
boil	v	2	4	5	4	3
book	n	1	1	1	1	1
border	n	2	4	6	4	6
born	v	2	3	1	1	2
borrow	v	3	4	2	2	5
bottle	n	1	1	1	1	1
box	n	1	1	1	1	1
boy	n	1	1	1	1	1
brave	adj	2	3	3	2	1
bread	n	1	1	3	1	1
break	v	2	2	1	1	3
breakfast	n	1	1	1	1	1
breathe	v	3	3	4	3	4
bridge	n	1	2	2	3	2
bring	v	1	2	1	1	1
brother	n	1	1	1	1	1
brush	n	2	3	3	3	3
bucket	n	1	4	5	3	5
build	v	2	1	2	1	2
building	n	1	1	1	1	3
bus	n	1	1	2	1	1
business	n	2	2	2	2	3
busy	adj	2	3	2	2	1
but	conj	1	1	1	1	1
butter	n	1	2	3	1	2
button	n	1	4	3	1	4
buy	v	1	1	2	1	1
call	v	2	2	1	1	1
calm	adj	2	5	3	2	4
camera	n	1	3	1	2	4
car	n	1	1	1	1	1
card	n	2	2	3	2	1
careful	adj	2	3	1	1	2
carry	v	1	2	1	1	2
cat	n	1	1	1	1	1
chair	n	1	1	1	1	1
chance	n	2	3	3	3	3
change	v	2	1	1	2	1
character	n	4	5	4	4	4
cheap	adj	1	2	1	1	3
cheat	v	4	3	3	3	5
check	v	4	4	3	3	5

Total: 140

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
cheque	n	3	3	3	3	4
chicken	n	1	3	2	2	3
chocolate	n	2	2	2	2	6
church	n	1	2	2	1	2
cinema	n	1	1	1	1	2
circle	n	2	3	2	2	3
city	n	1	2	4	1	2
clap	v	4	5	5	4	6
clear	adj	2	3	3	2	2
clerk	n	2	4	4	4	6
clever	adj	3	3	2	1	2
climb	v	3	2	3	1	2
clock	n	1	1	1	2	2
close	v	1	1	1	1	2
clothes	n	2	1	2	2	1
cloud	n	2	2	2	2	2
coast	n	2	3	3	2	3
coat	n	1	1	1	1	1
coffee	n	1	2	1	1	2
coin	n	4	3	3	3	5
college	n	4	3	4	4	6
colour	n	1	2	3	1	2
comb	n	3	3	4	3	5
comb	v	3	3	4	3	5
comfortable	adj	3	3	2	2	3
common	adj	3	4	5	3	3
compare	v	3	4	5	3	5
complain	v	4	5	5	3	4
condition	n	4	3	4	4	4
congratulate	v	4	5	5	3	6
connect	v	4	4	5	4	6
contain	v	3	3	3	3	4
continue	v	4	3	3	2	5
control	v	3	4	4	3	5
conversation	n	2	2	2	2	4
cook	n	1	2	2	1	2
cook	v	1	2	2	1	1
cool	adj	3	3	3	3	3
copy	n	2	3	4	3	4
copy	v	2	3	4	3	4
corner	n	2	2	2	1	3
correct	adj	3	5	5	3	4
cost	n	2	2	2	2	2
cough	n	2	4	6	3	5
count	v	2	3	2	2	4
cousin	n	2	3	4	2	3
cover	n	3	3	3	3	2

Total: 187

B4

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
cover	v	3	3	3	3	2
cow	n	1	2	2	3	1
coward	n	4	4	6	3	5
crash	n	3	3	2	2	4
crash	v	3	3	2	2	4
crime	n	3	3	2	3	5
crowd	n	2	3	2	2	3
cruel	adj	2	4	4	3	4
cup	n	1	2	2	1	1
curious	adj	3	5	5	4	3
curve	n	3	4	5	3	6
curve	v	3	4	5	3	6
customer	n	2	5	3	2	5
cut	v	1	2	2	1	2
damage	n	3	4	3	3	5
damp	adj	3	6	6	5	6
dance	n	1	2	2	1	2
dance	v	1	2	2	1	2
danger	n	3	2	2	1	2
dangerous	adj	3	2	2	1	2
dark	adj	2	2	1	1	2
daughter	n	1	1	1	1	1
day	n	1	1	1	1	1
dead	adj	2	1	1	1	1
deaf	adj	3	5	5	3	6
death	n	2	3	2	3	3
decide	v	3	3	5	2	3
deep	adj	2	2	2	2	2
defeat	n	4	4	4	4	6
defeat	v	4	4	4	4	6
delay	n	3	4	4	3	4
deliver	v	3	4	5	3	4
demand	v	4	4	5	4	4
describe	v	3	3	2	2	4
desert	n	2	3	5	2	4
deserve	v	4	6	6	4	4
desk	n	1	2	2	1	5
destroy	v	4	3	3	4	3
develop	v	4	4	4	4	6
diamond	n	2	4	4	3	6
die	v	2	1	1	1	1
different	adj	2	2	1	1	1
difficult	adj	1	2	1	1	3
dig	v	2	3	3	3	4
dinner	n	1	1	1	1	3
dirt	n	2	1	4	4	1
dirty	adj	1	1	1	1	1

Total: 234

B5

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
disappoint	v	4	4	4	4	6
disease	n	4	3	3	4	5
district	n	4	4	5	4	6
disturb	v	3	5	5	5	5
divide	v	4	3	5	4	4
do	v	1	1	1	1	1
doctor	n	1	2	1	1	1
dog	n	1	1	1	1	1
dollar	n	1	2	1	1	6
door	n	1	1	1	1	1
double	adj	2	3	4	3	3
down	prep	1	1	1	1	1
downstairs	adv	3	2	1	2	3
dream	n	2	3	3	3	3
dream	v	2	3	3	3	3
dress	n	1	2	1	1	1
dress	v	1	2	1	1	1
drink	n	2	1	1	1	1
drink	v	1	1	1	1	1
drive	v	1	1	1	1	2
drown	v	4	5	4	4	4
drum	n	2	4	5	5	5
dry	adj	2	1	3	2	2
dry	v	2	1	3	2	2
during	prep	3	2	2	2	5
each other	pron	4	3	3	3	2
ear	n	1	1	1	1	1
early	adv	1	2	1	1	2
earn	v	3	4	4	3	4
east	adj	2	2	1	1	2
east	adv	2	2	1	1	2
east	n	2	2	1	1	2
easy	adj	1	2	1	1	1
eat	v	1	1	1	1	1
edge	n	3	3	3	3	3
educate	v	4	4	5	4	6
effect	n	4	3	4	3	5
efficient	adj	4	4	5	4	6
egg	n	1	2	2	2	1
electric	adj	1	3	4	2	5
elephant	n	1	3	6	4	5
else	adv	4	2	4	4	3
empty	adj	2	2	2	2	3
end	n	1	1	3	2	2
end	v	2	1	3	2	4
enemy	n	4	5	2	2	3
engine	n	3	3	3	3	4

Total: 281

B6

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
enjoy	v	3	2	2	2	3
enough	adv	2	2	3	2	2
enter	v	2	3	3	3	3
entertain	v	4	5	6	3	4
enthusiastic	adj	3	6	4	3	4
envelope	n	1	4	5	1	6
escape	n	4	2	2	2	3
escape	v	4	2	2	2	3
evening	n	1	1	1	1	1
ever	adv	2	2	3	2	4
exist	v	4	5	6	4	6
expect	v	2	3	4	4	3
expensive	adj	1	2	1	1	2
experience	n	4	4	4	4	4
experiment	n	4	4	6	5	6
explain	v	2	3	2	2	3
explode	v	4	3	3	4	5
explore	v	4	4	3	4	5
extra	adj	3	5	3	3	5
eye	n	1	1	1	1	1
face	n	1	1	1	1	1
factory	n	1	4	2	1	5
fail	v	4	3	3	3	4
fall	v	2	2	2	2	1
false	adj	2	4	4	3	6
family	n	1	1	1	1	2
famous	adj	3	2	1	1	3
far	adv	1	1	1	1	1
farm	n	2	2	3	2	1
fat	adj	1	1	1	1	1
father	n	1	1	1	1	1
fault	n	2	4	4	4	4
favourite	adj	2	4	2	2	3
fear	n	4	3	3	3	4
feed	v	3	3	5	3	3
female	adj	3	5	6	2	6
fence	n	4	6	6	5	3
fever	n	3	5	4	4	6
fight	n	1	1	2	1	2
final	adj	4	6	3	2	5
find	v	1	1	1	1	1
fine	adj	1	3	2	2	2
finger	n	1	1	2	1	3
finish	v	1	2	1	1	2
fire	n	1	1	2	1	1
fish	n	1	1	2	1	1
fit	v	4	3	3	3	5

Total: 328

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
fix	v	3	3	3	3	4
flag	n	3	3	3	3	4
float	v	4	5	5	4	5
flood	n	4	4	5	5	6
floor	n	1	1	2	1	2
flower	n	1	1	1	1	1
fly	v	1	2	2	1	2
follow	v	2	2	2	1	2
food	n	1	1	2	1	1
football	n	1	1	1	2	2
for	prep	2	1	1	1	1
forbid	v	4	4	5	4	4
foreign	adj	2	4	3	2	4
forest	n	2	3	3	2	2
forget	v	1	2	1	1	3
forgive	v	4	4	5	2	5
fork	n	1	2	2	2	2
free	adj	2	2	1	2	2
freeze	v	3	5	4	3	5
fresh	adj	2	3	4	2	2
fridge	n	2	4	3	2	3
friend	n	1	1	1	1	1
from	prep	1	1	1	1	1
fruit	n	1	1	2	1	2
full	adj	2	2	2	2	1
funny	adj	2	3	2	2	3
furniture	n	2	3	3	2	3
game	n	1	1	3	1	1
garage	n	1	3	2	1	2
garden	n	1	1	1	1	1
gas	n	1	1	4	2	5
gate	n	4	4	4	3	3
gather	v	4	4	5	4	6
generous	adj	4	5	4	4	4
gentle	adj	4	3	3	3	2
girl	n	1	1	1	1	1
give	v	1	1	1	1	1
glad	adj	2	3	4	2	1
gold	n	2	2	2	2	1
good	adj	1	1	1	1	1
govern	v	4	4	5	4	6
grass	n	1	1	3	2	2
grateful	adj	4	5	3	3	5
great	adj	2	2	5	2	2
green	adj	1	1	1	1	1
grey	adj	1	1	2	2	2
group	n	3	3	2	2	4

Total: 375

B8

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
guard	n	2	4	4	2	5
guard	v	2	4	4	2	3
guess	v	3	4	3	2	3
guide	n	3	4	4	3	3
guide	v	3	4	4	3	3
hair	n	1	2	2	1	1
hammer	n	1	4	5	4	4
hand	n	1	1	1	1	1
handle	n	1	4	5	4	5
hang	v	3	3	5	3	3
happen	v	1	2	1	1	2
happy	adj	1	1	1	1	1
hard	adj	1	2	2	1	1
hardly	adv	3	4	6	4	4
harm	v	3	5	4	3	5
harvest	n	4	5	6	5	6
harvest	v	4	5	6	5	6
hat	n	1	2	1	1	1
hate	v	2	3	3	3	4
head	n	1	1	1	1	1
heal	v	4	5	6	5	6
hear	v	1	1	1	1	1
heart	n	2	2	3	2	2
heat	v	3	3	3	3	3
heaven	n	4	5	6	3	6
heavy	adj	1	2	2	1	2
help	v	1	1	1	1	1
here	adv	1	1	1	1	1
hide	v	4	2	2	2	1
high	adj	2	2	2	2	2
hill	n	2	2	1	2	1
hire	v	3	5	4	4	5
history	n	3	4	6	4	3
hit	v	1	2	1	1	1
hobby	n	2	4	4	3	3
hold	v	2	2	2	2	2
holiday	n	2	2	2	1	3
hollow	adj	4	6	5	3	5
home	n	1	1	1	1	1
honest	adj	3	4	3	2	4
hope	n	2	2	3	2	4
hope	v	2	2	3	2	2
horizon	n	4	6	6	5	4
horse	n	1	2	2	2	1
hospital	n	1	3	1	1	3
hot	adj	1	1	1	1	1
hotel	n	1	1	1	1	3

Total: 422

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
         NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
hour	n	1	2	1	1	1
house	n	1	1	1	1	1
human	adj	4	4	4	3	5
human	n	4	4	4	3	5
hungry	adj	1	3	1	1	2
hurry	v	2	3	2	2	3
hurt	v	2	2	2	2	1
husband	n	1	1	1	1	1
ice	n	2	3	3	2	3
idea	n	2	3	3	2	1
if	conj	2	2	2	2	1
ill	adj	2	1	1	1	1
imagine	v	3	4	4	3	4
important	adj	2	3	1	1	3
impossible	adj	3	3	3	3	5
improve	v	4	5	6	4	4
in	prep	1	1	1	1	1
in order to	conj	3	4	3	3	3
include	v	4	4	5	4	5
increase	v	4	4	6	4	6
indeed	adv	4	4	3	4	3
independent	adj	4	4	5	3	5
information	n	2	4	3	2	6
insect	n	3	4	4	3	6
inside	adv	2	2	2	1	3
insult	v	3	6	6	3	6
interest	v	4	3	3	3	4
interesting	adj	2	3	1	1	3
interfere	v	4	5	6	4	6
international	adj	3	4	4	3	4
interrupt	v	4	6	6	4	6
into	prep	1	1	1	1	1
introduce	v	4	5	3	3	4
invent	v	3	4	5	4	5
invitation	n	2	4	3	3	5
invite	v	3	4	3	3	3
island	n	1	2	2	2	2
jacket	n	1	4	5	1	3
jealous	adj	4	5	4	3	6
jeans	n	1	3	4	1	5
job	n	1	2	2	1	2
join	v	4	3	3	3	3
joke	n	3	3	5	3	5
journey	n	2	3	2	2	1
judge	n	4	4	4	4	3
juice	n	2	4	5	2	3
jump	v	1	1	2	1	1

Total: 469

B10

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
just	adv	2	2	2	2	1
keep	v	3	2	2	2	2
key	n	1	3	2	1	1
kick	v	1	2	4	1	3
kill	v	1	1	1	1	1
kind	adj	2	2	2	2	1
kind	n	2	2	3	2	3
king	n	2	2	4	2	2
kiss	n	2	2	4	2	3
kiss	v	2	2	4	2	3
kitchen	n	1	2	2	1	1
knee	n	2	4	4	2	3
knife	n	1	2	1	1	2
knock	v	2	4	2	2	4
know	v	1	2	1	1	1
knowledge	n	4	4	6	5	4
ladder	n	2	5	4	2	5
lake	n	2	3	3	2	3
lamp	n	1	4	5	2	1
language	n	2	6	5	1	2
large	adj	1	2	2	2	1
last	adj	2	1	1	1	1
last	adv	2	2	5	1	2
late	adj	1	2	1	1	1
laugh	v	2	2	1	1	1
lay	v	4	4	5	4	6
lazy	adj	3	4	2	2	1
leaf	n	2	3	2	2	2
learn	v	1	1	1	1	2
leather	adj	2	4	5	2	5
leather	n	2	4	5	2	5
leave	v	2	2	1	1	1
leg	n	1	1	1	1	1
lend	v	3	4	3	3	5
less	adv	2	3	3	3	2
library	n	2	3	3	2	3
licence	n	4	6	6	4	5
life	n	2	1	2	3	2
lift	n	3	5	3	2	3
lift	v	4	3	3	2	3
like	v	1	1	1	1	1
line	n	1	1	5	1	1
list	n	2	3	4	2	2
little	adj	1	1	1	1	1
live	v	1	1	1	1	1
load	n	3	4	6	5	2
load	v	3	4	5	5	4

Total: 516

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
         NB Usual caveats.   Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
lonely	adj	4	3	2	3	5
long	adj	1	1	1	1	1
look	v	1	1	2	1	1
lorry	n	1	2	5	2	2
loud	adj	2	2	2	2	2
love	v	1	1	1	1	1
low	adj	2	2	3	2	2
luck	n	2	3	2	1	4
lunch	n	1	2	2	1	3
main	adj	4	4	4	4	5
m e	v	2	1	1	1	1
male	adj	3	5	6	2	6
map	n	3	3	3	3	3
mark	n	3	3	6	3	2
mark	v	3	3	6	3	2
market	n	1	2	2	1	2
marry	v	3	2	2	2	1
meal	n	2	1	2	2	2
measure	v	4	4	3	3	2
meat	n	1	3	1	1	1
medicine	n	3	3	3	3	5
meet	v	2	2	1	1	2
melt	v	3	4	5	5	5
mend	v	3	4	2	2	3
message	n	4	4	3	3	4
middle	n	2	2	2	2	1
mile	n	1	3	1	1	2
milk	n	1	2	1	1	1
mind	n	4	3	6	3	5
m ute	n	1	2	1	1	2
Miss	title	1	1	1	1	2
mistake	n	3	3	2	1	2
modern	adj	3	2	2	2	4
modest	adj	4	6	6	4	6
moment	n	1	3	3	3	2
money	n	1	1	1	1	1
monkey	n	1	4	5	4	5
month	n	1	1	1	1	1
mood	n	3	4	5	3	4
moon	n	1	2	2	2	2
morning	n	1	1	1	1	1
mother	n	1	1	1	1	1
mountain	n	1	1	2	1	2
mouth	n	1	1	1	1	1
move	v	2	2	1	1	2
murder	n	3	3	2	2	5
murder	v	3	3	2	2	5

Total: 563

Word	Category	LSR	CEL	OUP	PEN	MAC
music	n	1	2	2	2	2
mystery	n	3	3	2	2	4
name	n	1	1	1	1	1
narrow	adj	2	4	2	2	2
natural	adj	3	3	3	2	4
nature	n	4	3	4	4	4
near	prep	1	1	1	1	1
nearly	adv	2	3	1	2	1
necessary	adj	2	3	2	3	3
neck	n	1	3	2	2	3
need	v	2	3	1	2	3
nephew	n	4	4	5	3	5
nervous	adj	2	5	6	3	5
net	n	4	5	6	4	6
never	adv	1	1	1	1	1
new	adj	1	1	1	1	1
newspaper	n	1	2	2	2	2
next	adj	1	2	1	1	1
next	adv	2	3	1	1	1
nice	adj	1	3	1	1	1
niece	n	4	4	5	3	5
night	n	1	1	1	1	1
nobody	pron	3	3	1	1	2
noise	n	2	2	1	2	1
none	pron	3	3	2	2	1
nonsense	n	4	4	5	4	5
north	adj	2	2	1	1	1
north	adv	2	2	1	1	1
nose	n	1	1	2	1	1
not	adv	1	1	1	1	1
note	n	3	2	2	3	5
nothing	pron	1	1	1	1	1
notice	n	3	3	3	4	3
notice	v	3	3	3	2	3
now	adv	1	1	1	1	1
number	n	1	1	1	1	2
nurse	n	1	3	1	2	5
o'clock	adv	1	1	1	1	1
object	n	4	4	6	4	3
observe	v	3	6	6	4	4
of	prep	1	1	1	1	1
off	prep	1	1	1	1	2
often	adv	1	1	1	1	1
old	adj	1	1	1	1	1
on	prep	1	1	1	1	1
only	adj	1	2	2	1	2
open	adj	1	1	1	1	1

Total: 610

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
         NB Usual caveats.   Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
open	v	1	1	1	1	1
opportunity	n	3	6	6	3	5
opposite	adj	4	4	2	2	4
or	conj	1	1	1	1	1
orange	adj	1	1	1	1	3
orange	n	1	1	1	1	3
overtake	v	4	5	6	4	5
owe	v	3	4	4	4	4
own	v	3	3	3	3	3
packet	n	1	4	3	2	5
page	n	2	1	2	2	2
pain	n	2	4	3	3	2
paint	n	3	3	4	2	3
paint	v	3	3	4	2	3
palace	n	2	3	3	3	5
paper	n	1	2	2	1	1
park	n	1	3	3	1	3
part	n	2	2	5	2	1
part	v	2	5	6	5	1
party	n	3	2	2	2	3
pass	v	2	2	3	3	2
passenger	n	2	3	2	2	5
patience	n	4	5	6	4	6
patient	adj	3	5	3	3	5
pattern	n	3	4	4	5	6
pause	n	3	4	4	3	5
pause	v	3	4	4	3	5
pay	v	1	2	2	1	3
peace	n	4	3	3	3	3
pen	n	1	1	1	1	1
pencil	n	1	1	2	1	1
people	n	1	1	1	1	1
perfect	adj	2	4	4	2	3
perhaps	adv	2	1	1	1	2
permission	n	4	5	5	4	6
person	n	2	1	2	2	1
pet	n	2	4	5	3	6
petrol	n	3	3	5	3	5
pick up	v	3	3	3	2	3
picture	n	1	1	1	1	1
pig	n	1	5	6	2	3
pile	n	3	4	5	3	4
pink	adj	2	3	3	2	6
pipe	n	1	3	3	2	1
pity	n	4	4	6	2	4
place	n	2	1	2	1	1
plan	n	3	3	2	2	3

Total: 657

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
         NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
plant	n	2	2	3	2	1
plastic	adj	2	4	5	2	3
plate	n	1	2	2	1	1
play	v	1	1	1	1	1
pleasant	adj	4	4	5	2	3
please	v	1	4	1	4	1
pleasure	n	4	4	5	4	5
pocket	n	2	2	1	2	3
poem	n	4	5	6	4	6
point	n	2	2	5	4	2
prison	n	4	4	5	3	4
polite	adj	3	4	3	2	2
poor	adj	2	1	2	2	1
popular	adj	3	3	3	3	5
port	n	3	3	6	3	5
position	n	4	4	5	4	5
possess	v	3	5	6	5	4
possible	adj	2	2	2	2	3
postpone	v	3	6	6	4	6
pot	n	2	3	3	2	1
potato	n	1	2	2	2	3
pour	v	3	4	5	3	2
power	n	4	3	6	4	2
practice	n	3	4	5	2	3
practise	v	3	4	5	2	4
prefer	v	3	4	2	2	5
prepare	v	4	4	4	3	5
pretty	adj	1	2	3	1	1
prevent	v	4	4	4	4	6
price	n	1	2	3	2	3
prison	n	3	3	2	2	3
private	adj	3	4	4	3	4
prize	n	3	4	5	3	6
probable	adj	4	3	3	4	4
problem	n	2	4	3	2	4
proof	n	4	4	4	4	5
protect	v	4	6	6	4	4
prove	v	4	3	4	4	3
public	adj	3	4	5	3	3
public	n	3	4	5	3	3
pull	v	1	3	2	1	2
punish	v	4	4	3	3	5
pure	adj	4	4	6	3	4
push	v	1	3	2	1	2
put	v	1	1	1	1	1
quality	n	3	4	5	3	5
quantity	n	3	4	6	3	5

Total: 704

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
quarrel	v	4	4	6	4	1
quarter	n	2	2	1	1	3
queen	n	2	2	4	2	2
question	n	1	1	1	1	2
queue	n	3	5	4	4	3
queue	v	3	5	4	4	3
quick	adj	2	3	1	1	1
quiet	adj	1	2	1	1	2
radio	n	1	1	2	1	2
railway	n	2	3	3	2	2
rain	n	1	1	1	1	1
rain	v	1	1	1	1	1
raise	v	3	4	5	5	2
rather	adv	3	4	4	4	1
raw	adj	2	5	6	3	6
razor	n	2	4	6	3	6
read	v	1	1	1	1	1
ready	adj	2	1	1	1	1
real	adj	2	2	2	2	2
reason	n	4	3	3	4	2
receive	v	4	3	4	3	2
recommend	v	4	6	5	4	6
recover	v	4	5	6	4	5
red	adj	1	1	1	1	1
refuse	v	4	3	3	3	4
regret	v	4	5	6	4	5
religion	n	4	5	6	4	5
remain	v	4	3	4	4	2
remember	v	1	2	1	1	2
remind	v	4	4	4	3	5
repair	v	3	4	4	3	5
repeat	v	2	3	3	2	4
reply	v	4	4	2	2	3
report	v	4	3	3	2	3
rescue	v	4	4	5	4	5
respect	n	4	6	6	4	5
respect	v	4	6	6	4	5
rest	n	2	3	4	2	2
rest	v	2	3	2	2	2
restaurant	n	1	2	2	1	3
result	n	3	3	4	3	4
reward	n	3	5	6	3	6
rice	n	1	4	2	2	4
rich	adj	2	1	1	1	1
ride	v	2	2	2	2	1
ring	n	3	3	2	2	1
rise	v	3	3	4	4	3

Total: 751

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
         NB Usual caveats.   Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
river	n	1	1	1	1	1
road	n	1	1	1	1	1
rock	n	3	3	3	3	2
roof	n	1	3	3	2	2
room	n	1	1	1	1	1
rough	adj	4	3	4	2	4
round	adj	2	1	2	1	2
rubber	n	3	3	5	2	5
rude	adj	4	4	5	2	5
run	v	1	1	1	1	1
rush	v	3	4	5	3	4
s	adj	1	1	3	1	1
safe	adj	3	3	3	2	2
safety	n	4	4	5	4	4
sail	n	3	3	2	2	6
sail	v	3	3	2	2	6
sake	n	4	4	6	4	6
sale	n	3	5	5	3	6
salt	n	1	4	2	2	4
sand	n	2	2	5	4	2
sandwich	n	1	5	1	3	3
satisfy	v	4	5	6	3	4
saucer	n	1	6	4	3	6
say	v	1	1	1	1	1
school	n	1	1	2	1	2
science	n	4	1	5	3	4
sea	n	3	1	1	1	1
search	n	4	3	3	2	4
season	n	4	4	4	2	3
seat	n	3	2	2	3	3
second	n	1	2	1	2	3
secret	n	3	3	2	2	4
seed	n	3	3	6	3	6
seem	v	3	2	3	2	2
sell	v	2	1	2	1	1
send	v	2	1	2	1	1
sense	n	4	3	5	3	5
separate	adj	4	4	4	3	4
separate	v	4	4	4	3	3
service	n	4	5	5	4	3
several	pron	3	3	4	2	2
shadow	n	4	3	3	2	3
shake	v	4	3	2	2	4
shall	v	3	2	1	2	2
shallow	adj	2	3	3	2	5
shape	n	2	2	3	2	2
sharp	adj	3	4	4	5	4

Total: 798

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
         NB Usual caveats.   Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
e--v-	-------2-	-4-	-5-	-2-	-6-	---
sheep	n	3	4	2	2	4
shine	v	1	2	3	2	3
ship	n	1	1	1	1	3
shirt	n	1	1	1	1	2
shoot	v	2	2	2	2	5
short	adj	1	1	2	1	1
shout	v	2	1	2	2	3
show	v	2	2	2	1	1
shut	v	1	1	3	1	1
side	n	2	2	2	2	1
sign	n	3	3	4	3	4
silent	adj	4	3	2	2	3
silly	adj	3	4	5	3	2
silver	adj	2	3	2	2	2
silver	n	2	3	2	2	2
sing	v	2	3	1	1	1
sink	v	4	4	6	4	5
sister	n	1	1	1	1	1
sit	v	1	1	1	1	1
size	n	2	2	6	2	3
skin	n	3	3	3	2	3
skirt	n	1	2	1	1	3
sky	n	1	1	1	1	1
slow	adj	1	1	1	1	2
small	adj	1	1	2	1	1
smell	n	2	2	2	2	1
smell	v	2	2	2	2	1
smile	n	2	2	1	1	2
smile	v	2	2	1	1	2
smoke	n	2	2	1	1	3
smoke	v	1	2	1	1	3
smooth	adj	4	4	4	2	5
snake	n	2	4	6	2	6
snow	n	1	2	2	2	2
snow	v	1	2	2	2	2
soap	n	1	2	4	2	3
soft	adj	1	2	2	2	2
solve	v	3	4	5	3	6
some	pron	1	1	3	2	1
sometimes	adv	2	1	1	1	1
son	n	1	1	1	1	1
song	n	3	3	2	2	1
soon	adv	2	1	1	1	1
sore	adj	4	5	5	4	6
sorry	adj	1	1	1	1	1
sound	n	2	2	2	1	2

Total: 845

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
soup	n	1	3	4	1	3
sour	adj	2	4	6	3	6
south	adj	2	2	1	1	1
south	adv	2	2	1	1	1
south	n	2	2	1	2	1
speak	v	1	1	2	1	1
special	adj	3	4	2	2	4
speech	n	3	5	6	4	5
spend	v	2	2	2	2	3
spoon	n	1	2	2	2	2
sport	n	1	2	3	2	2
spring	n	2	2	1	1	2
square	adj	2	3	2	1	4
square	n	2	3	2	1	4
stamp	n	1	3	2	2	2
stand	v	1	1	1	1	1
start	v	1	2	3	1	2
station	n	1	3	2	1	1
stay	v	2	2	1	1	1
steal	v	2	3	2	2	4
steel	n	2	3	6	4	5
step	n	3	3	3	3	2
still	adv	2	1	2	2	1
stomach	n	1	5	2	2	2
stone	n	1	2	2	2	1
stop	v	1	1	1	1	2
storm	n	3	2	2	2	3
story	n	1	1	1	1	1
straight	adj	2	3	3	2	2
strange	adj	3	2	2	2	2
street	n	1	1	5	1	1
strong	adj	1	2	2	1	2
study	v	3	3	3	2	3
stupid	adj	2	5	2	2	5
sugar	n	1	3	3	1	1
suggest	v	4	4	5	3	5
suit	n	1	3	1	2	4
summer	n	2	2	1	1	2
sun	n	1	1	3	1	1
suppose	v	4	2	3	3	3
sure	adj	2	3	2	2	2
surprise	n	3	3	2	2	2
surprise	v	3	3	2	2	2
suspect	v	3	4	4	3	6
sweet	adj	2	2	2	2	2
swim	v	1	2	2	1	2
switch	n	3	4	6	2	3

Total: 892

R19

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
syste	n	4	6	6	4	6
table	n	1	1	1	1	1
tail	n	3	5	3	3	4
take	v	1	1	1	1	1
talk	v	2	1	1	1	2
tall	adj	1	1	1	1	1
tame	adj	4	5	6	4	6
taste	n	2	2	2	2	1
tax	n	3	5	6	5	6
taxi	n	1	1	1	1	2
teach	v	1	1	1	2	2
team	n	3	3	4	3	3
t ephone	n	1	1	1	1	2
tell	v	1	1	1	1	1
temperature	n	3	4	4	3	5
tent	n	2	4	5	2	6
terrible	adj	2	3	2	2	4
test	n	3	4	4	3	5
test	v	4	4	4	3	5
than	conj	2	2	2	1	2
that	conj	1	2	2	1	2
the	det	1	1	1	1	1
theatre	n	2	4	2	2	5
thick	adj	1	3	3	1	2
thief	n	2	2	2	2	5
thin	adj	1	2	2	1	2
thing	n	1	1	1	1	1
think	v	1	1	1	1	1
through	prep	2	1	1	1	1
throw	v	1	3	4	1	1
ticket	n	1	1	1	1	2
t.	n	3	3	4	2	3
tie	v	3	3	4	2	3
till	conj	4	3	5	4	2
time	n	1	1	1	1	1
timetable	n	4	4	5	2	5
tin	n	3	4	5	1	1
to	prep	1	1	1	1	1
today	adv	1	1	1	1	1
together	adv	3	2	2	1	2
tomorrow	adv	1	1	1	1	1
tongue	n	1	3	3	2	2
tonight	adv	1	1	1	1	1
touch	v	2	1	2	2	3
towel	n	3	4	5	2	3
town	n	1	1	1	1	1
tractor	n	2	5	6	5	6

Total: 939

B20

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
train	n	1	1	1	1	1
train	v	1	5	4	4	4
travel	n	2	3	2	2	3
travel	v	2	3	2	2	3
tree	n	1	1	1	1	1
trial	n	4	4	4	4	6
trip	n	3	4	4	3	3
trousers	n	1	2	1	1	2
true	adj	2	1	1	2	2
trust	n	4	4	6	4	3
trust	v	4	4	6	4	3
try	v	2	2	2	1	2
t n	v	1	2	2	1	2
twice	adv	2	2	2	2	1
type	n	3	3	3	2	5
ugly	adj	1	4	4	2	1
umbrella	n	1	4	3	2	3
uncle	n	2	3	4	2	3
under	prep	1	1	1	1	2
understand	v	1	2	1	1	1
university	n	1	3	3	2	5
unless	conj	3	4	4	3	4
until	prep	3	3	2	2	2
up	prep	1	1	1	1	1
upstairs	adv	3	3	1	2	3
valley	n	2	3	3	2	4
value	n	4	3	6	3	4
very	adv	1	1	1	1	1
view	n	3	4	4	3	4
visit	v	2	2	1	1	2
voice	n	2	2	2	2	2
v e	n	4	4	6	4	6
vote	v	4	4	6	4	6
voyage	n	3	4	4	3	5
walk	v	1	1	1	1	1
wall	n	1	2	2	2	1
warm	adj	1	2	1	1	2
warn	v	4	4	3	2	5
waste	v	3	3	6	2	2
water	n	1	1	1	1	1
wave	v	2	2	4	2	2
weak	adj	1	3	5	1	3
wear	v	3	2	1	1	2
weather	n	1	1	1	1	3
weight	n	3	4	6	3	3
west	n	2	2	1	1	2
wet	adj	2	1	1	1	2

Total: 986

B21

Query: What words are in LSR, CEL, OUP, PEN, MAC and at what level?
        NB Usual caveats. Data source: CUPRAW

Word	Category	LSR	CEL	OUP	PEN	MAC
wheel	n	2	1	2	2	2
white	adj	1	1	1	1	1
wide	adj	2	2	4	2	2
wife	n	1	2	1	1	1
wild	adj	4	3	2	2	3
win	v	1	2	3	2	2
wind	n	1	2	2	1	1
window	n	1	1	1	1	1
wine	n	1	4	5	2	3
wing	n	2	5	2	3	3
winter	n	2	2	1	1	2
wipe	v	4	6	6	4	6
wish	n	3	3	3	3	2
wish	v	3	3	3	3	2
with	prep	1	1	1	1	1
without	prep	1	1	1	1	2
wonder	v	3	3	3	3	6
wonderful	adj	2	3	1	2	3
word	n	1	1	2	1	2
work	n	1	1	1	1	1
work	v	1	1	1	1	1
world	n	1	1	2	1	2
worry	v	2	3	3	2	5
worse	adj	2	2	1	1	2
worst	adj	2	2	1	1	3
worst	adv	2	2	1	1	3
wound	n	4	4	4	4	6
wound	v	4	4	4	4	6
wrap	v	3	5	5	4	6
wreck	n	3	3	6	3	6
write	v	1	1	1	1	1
wrong	adj	1	1	1	1	2
year	n	1	1	1	1	1
yellow	adj	1	1	1	1	2
yes	coll	1	1	1	1	1
yet	adv	2	2	2	2	1
young	adj	1	1	1	1	1

Total: 1023

B22

APPENDIX "C"

Query: What words occur in Threshold, but NOT in LSR, CEL, OUP, PEN
        or MAC? Data Source: CUPRAW. (Therefore usual caveats:
        spelling of Business-man etc)

Word	Category	THR
addict	n	0
Africa	n	0
ahem!	coll	0
airforce	n	0
allowed	adj	0
ally	n	0
America	n	0
Asia	n	0
Australia	n	0
b gage	n	0
ball-pen	n	0
ball-point	adj	0
ballet	n	0
bank-note	n	0
biography	n	0
biro	n	0
boarding-pass	n	0
booking office	n	0
bookshop	n	0
booth	n	0
boy-friend	n	0
bureau	n	0
burglary	n	0
business-man	n	0
business-woman	n	0
bye-bye	coll	0
cabaret	n	0
cabin	n	0
c l-box	n	0
canteen	n	0
carpentry	n	0
cassette recorder	n	0
cauliflower	n	0
CD	abbr	0
chair (person)	n	0
charter	n	0
check in	v	0
check out	v	0
cheerio	coll	0
chips	n	0
closing-time	n	0
cod	n	0
coffee-shop	n	0
come and see	v	0
Common Market	n	0
compact	adj	0

Total: 46

C1.

Query: What words occur in Threshold, but NOT in LSR, CEL, OUP, PEN
         or MAC? Data Source: CUPRAW. (Therefore usual caveats:
         spelling of Business-man etc)

Word	Category	THR
cricket	n	0
cyclist	n	0
dancer	n	0
disarmament	n	0
disc jockey	n	0
▶ disco(theque)	n	0
dizzy	adj	0
documentary	adj	0
dressed	adj	0
dressing	n	0
drugs	n	0
E!	abbr	0
employer	n	0
faithfully	adv	0
fancy (that)!	coll	0
▶ fax	n	0
floor-show	n	0
fresh(ly)	adj	0
fruit-juice	n	0
furnished	adj	0
gale	n	0
gardening	n	0
girl-friend	n	0
go for	v	0
grandchild	n	0
greatly	adv	0
hallo/hello	coll	0
help-yourself	n	0
hi-jack	v	0
hi-jacker	n	0
hockey	n	0
hostel	n	0
in brief	adj	0
included	adj	0
incorrect	adj	0
interpret	v	0
jewelry	n	0
killed	adj	0
killer	n	0
kindly	adv	0
knitting	n	0
labourer	n	0
lawn	n	0
leisure	n	0
level crossing	n	0
locker	n	0
lover	n	0

Total: 93

C2

Query: What words occur in Threshold, but NOT in LSR, CEL, OUP, PEN
         or MAC? Data Source: CUPRAW. (Therefore usual caveats:
         spelling of Business-man etc)

Word	Category	THR
moist	adj	0
motor cycle	n	0
mug	n	0
musical	n	0
mutton	n	0
NATO	abbr	0
night-club	n	0
note-paper	n	0
odour	n	0
oh dear!	coll	0
oh!	coll	0
one-way	adj	0
peanut	n	0
pear	n	0
pen-friend	n	0
permitted	adj	0
photography	n	0
physician	n	0
plaice	n	0
player	n	0
playing-cards	n	0
policewoman	n	0
pollution	n	0
pop(ular)	adj	0
possibly	adv	0
poste-restante	adv	0
poultry	n	0
President	n	0
prime	adj	0
Prime Minister	n	0
put through	v	0
quay	n	0
racing	n	0
rainy	adj	0
raisin	n	0
rape	n	0
reading	n	0
recorder	n	0
refreshments	n	0
register	n	0
registration	n	0
revise	v	0
roll (bread)	n	0
rugby	n	0
ruins	n	0
saleswoman	n	0
sanitary	adj	0

Total: 140

C3

Query: What words occur in Threshold, but NOT in LSR, CEL, OUP, PEN
         or MAC? Data Source: CUPRAW. (Therefore usual caveats:
         spelling of Business-man etc)

Word	Category	THR
scooter	n	0
sculpture	n	0
self-service	n	0
sender	n	0
sh!	coll	0
shop assistant	n	0
smoking	n	0
snooker	n	0
snowstorm	n	0
so-so	adj	0
s ialist	n	0
sock	n	0
sole	n	0
spinach	n	0
steering	n	0
stocking	n	0
successful	adj	0
sum up	v	0
supposed to	adj	0
surprising	adj	0
tart	n	0
thingummyjig	n	0
third-party	adj	0
ton(ne)	n	0
tradesman	n	0
traveller	n	0
turkey	n	0
tut-tut	coll	0
ugh!	coll	0
u huh	coll	0
unable	adj	0
undressed	adj	0
unfortunately	adv	0
unhappy	adj	0
unimportant	adj	0
unintelligent	adj	0
unleaded	adj	0
unpleasant	adj	0
unsuccessful	adj	0
vanilla	n	0
veal	n	0
video recorder	n	0
videotape	n	0
W.C.	n	0
waiting-room	n	0
walkman	n	0
ward	n	0

Total: 187

C4

Query: What words occur in Threshold, but NOT in LSR, CEL, OUP, PEN
        or MAC? Data Source: CUPRAW. (Therefore usual caveats:
        spelling of Business-man etc)

Word	Category	THR
washing machine	n	0
well done!	coll	0
were	v	0
you know	int	0

Total: 191

C5

How are entries with "day" handled by each of the databases?

Word	LSR	CEL	OUP	PEN	MAC	WAY	THR	CHA	CEC
all day	2								
birthday	2	3	2	3	3		0	2	1
day	1	1	1	1	1	0			1
every day								1	1
everyday									3
Friday	1		1			0		1	
Happy Birthday									1
have a holiday		2							
holiday	2	2	2	1	3			2	1
holiday: on holiday									2
midday				2				1	4
M day	1		1			0		1	1
on holiday									1
package holiday								1	
Saturday	1		1			0		1	1
Sunday	1		1			0		1	1
the day before yesterday									1
Thursday	1		1			0		1	1
today	1	1	1	1	1	0		1	1
Tuesday	1		1			0		1	1
Wednesday			1			0		1	1
weekday						0			3
what a nice day		4							
yesterday		1	1	1	3	0		2	1

D.

NB This list needs to be taken with a grain of salt: some words
    are in other lists with (slightly) different spellings.

Word	Category	Level
air force	n	4
amazed	adj	4
amusement	n	4
anchor	n	2
arena	n	3
arrangement	n	4
arrogant	adj	4
ballpoint	n	1
biologist	n	4
brakes	n	3
budget	n	4
bus driver	n	1
camping site	n	2
classroom	n	1
coal mine	n	4
compliment	n	4
congratulations	n	4
cooperation	n	4
cowboy	n	1
daylight	n	3
daytime	n	3
depends on	v	3
digestion	n	4
diplomat	n	4
fan	n	3
flatter	v	4
flute	n	3
forge	v	4
forger	n	4
gas cooker	n	2
handlebar	n	2
handwriting	n	3
happily	adv	4
hinge	n	4
hoodlum	n	4
immigrate	v	4
information desk	n	2
introduction	n	4
invasion	n	4
Islam	n	4
Jordanian	adj	
junk	n	3
kangaroo	n	1
laser	n	4
leopard	n	1
lightning (written lightening)	n	3
litter	n	4

NB This list needs to be taken with a grain of salt: some words
     are in other lists with (slightly) different spellings.

Word	Category	Level
machinery	n	4
margarine	n	1
mass (religious)	n	4
micro-chip	n	4
mike (microphone)	n	4
mosque	n	4
movies	n	2
mugging	n	4
nap	n	3
nervously	adv	3
New Zealander	n	
news bulletin	n	4
pick up	v	4
passer-by	n	4
patiently	adv	3
peas	n	1
pedal	n	1
pirate	n	2
pocket-money	n	4
psychologist	n	4
radar	n	4
reduction	n	1
rub out	v	4
runway	n	2
rush hour	n	3
salute	n	3
shake one's head	v	4
sheriff	n	1
shoes	n	1
shoplifter	n	3
shoulders	n	2
specialize	v	4
stink	n	3
stormy	adj	3
take in (deceive)	v	4
teacup	n	2
telegramme	n	1
telephone number	n	1
toothbrush/paste	n	2
travel agency	n	4
virus	n	4
vitamin	n	3
washable	adj	4
washing-up	n	2
work/doesn't	v	2
worn out	adj	3
wrestle	v	3
wristwatch	n	2

E2

NB This list needs to be taken with a grain of salt: some words
    are in other lists with (slightly) different spellings.

Word	Category	Level
x-ray	n	3
yoghurt	n	1
zebra	n	1

E3