



The University of
Nottingham



The Knowledge-based List of Vocabulary (KVL): **Spanish Results**

Norbert Schmitt, Barry O'Sullivan, Laurence
Anthony, Karen Dunn, Benjamin Kremmel

The KVL

- A List of the Probabilities of Knowing Target Words
- “Knowledge-based Vocabulary List” (**KVL**)

Need for the KVL

Rank these items by learner knowledge:

- ___ Birthday
- ___ Cheat
- ___ Massive
- ___ Obviously
- ___ Pencil
- ___ Police

Need for the KVL

Rank these items by learner knowledge:

Frequency in SUBTlex

<u>4</u> Birthday	61
<u>5</u> Cheat	15
<u>3</u> Massive	108
<u>2</u> Obviously	216
<u>6</u> Pencil	11
<u>1</u> Police	286

Need for the KVL

- Frequency lists have been used to grade the ‘difficulty’ of vocabulary
- Works in a crude, overall way
- But many individual items are **misplaced**
- Because:
 - Learners do not get the amount of exposure indicated by corpus frequency figures
 - Corpus composition massively affects frequencies


Solution

- Compile a list of the probabilities of learners knowing the best-known 5,000 lemmas of English!
- In which order do learners learn words?

Purpose of the List

- To inform pedagogy, materials design, and test development
- Would lead to better selection of words on vocabulary size tests

Project Schedule

- **Year 1:** Conceptualization and development of tests
- **Year 2:** Tests developed (**7,679** lemmas), web site developed, and test site made live
- **Year 3:** Administration of the tests, analyses, and list development
 - Spanish: **150+** responses per item 
 - Chinese: **125+** responses per item
 - German: **60+** responses per item
- **Year 4:** Write book and article, and release of lists

Form-recall Test Format (Spanish)

casa Vivo en una casa grande que tiene tres
dormitorios.

h _ _ _ _

*house I live in a large house that has three
bedrooms.*

Form-recall Test Format (Spanish)

casa Vivo en una casa grande que tiene tres dormitorioses.

h o u s e

house I live in a large house that has three bedrooms.

Form-recall Test Format

- L1 → L2 1. *house* c a s a
- One correct answer
- Guessing is negligible
- Automatic computer scoring
- Practical

Test Format

- **Form-recall** level of mastery
- Must be **exact** spelling
- Test will set the bar at a relatively high level of mastery
- Correct answers on the test should ensure knowledge of the form-meaning link

Spanish Results (Best-known 50 lemmas)

1-10	11-20	21-30	31-40	41-50
dance	two	please	eat	atom
jet	one	family	shoe	west
city	human	three	my	table
zoo	hour	acid	beach	car
orange	winner	boy	clinic	tired
sun	hero	sing	play	air
German	book	drive	ice	red
day	change	atomic	key	class
no	fruit	federal	religion	bird
radio	love	black	hot	bye

Spanish Results (Best-known 50 lemmas)

1-10	11-20	21-30	31-40	41-50
dance	two	please	eat	atom
jet	one	family	shoe	west
city	human	three	my	table
zoo	hour	acid	beach	car
orange	winner	boy	clinic	tired
sun	hero	sing	play	air
German	book	drive	ice	red
day	change	atomic	key	class
no	fruit	federal	religion	bird
radio	love	black	hot	bye

Spanish Results (Best-known 50 lemmas)

1-10	11-20	21-30	31-40	41-50
dance	two	please	eat	atom
jet	one	family	shoe	west
city	human	three	my	table
zoo	hour	acid	beach	car
orange	winner	boy	clinic	tired
sun	hero	sing	play	air
German	book	drive	ice	red
day	change	atomic	key	class
no	fruit	federal	religion	bird
radio	love	black	hot	bye

Spanish Results (Best-known 50 lemmas)

- Difficulty was from about -5.00 to -4.00 on a Rasch scale
 - *dance* (-5.199)
 - *bye* (-4.022)
- Should we list individual lemmas?
 - Not much difference in degree of knowledge between adjacent lemmas
 - #10 *radio* (-4.511) vs. #11 *two* (-4.502)

Spanish Results (**Frequency rank**)

1-10		11-20		21-30		31-40		41-50	
dance	1,974	two	80	please	1,168	eat	543	atom	5,238
jet	3,163	one	51	family	147	shoe	1,427	west	665
city	290	human	399	three	135	my	44	table	539
zoo	6,371	hour	272	acid	3,259	beach	1,947	car	289
orange	3,296	winner	1,926	boy	382	clinic	2,404	tired	1,969
sun	1,237	hero	1,922	sing	1,014	play	200	air	370
German	1,833	book	241	drive	490	ice	1,234	red	598
day	90	change	356	atomic	5,498	key	1,468	class	431
no	93	fruit	1,668	federal	493	religion	1,387	bird	1,123
radio	905	love	579	black	253	hot	721	bye	19,365

23 within 500 | 27 better known than freq

1-10	11-20	21-30	31-40	41-50
dance 1,974	two 80	please 1,168	eat 543	atom 5,238
jet 3,163	one 51	family 147	shoe 1,427	west 665
city 290	human 399	three 135	my 44	table 539
zoo 6,371	hour 272	acid 3,259	beach 1,947	car 289
orange 3,296	winner 1,926	boy 382	clinic 2,404	tired 1,969
sun 1,237	hero 1,922	sing 1,014	play 200	air 370
German 1,833	book 241	drive 490	ice 1,234	red 598
day 90	change 356	atomic 5,498	key 1,468	class 431
no 93	fruit 1,668	federal 493	religion 1,387	bird 1,123
radio 905	love 579	black 253	hot 721	bye 19,365

Spanish Results (2,000-2,049)

2,000-09		2,010-19		2,020-29		2,030-39		2,040-49	
donate	4,127	decode	14,250	mentally	4,404	whale	4,267	economically	4,931
chain	1,815	cute	4,383	hunt	5,019	variety	1,172	bean	2,367
return	472	rider	3,854	caller	5,089	coach	1,034	really	143
ideally	7,997	success	777	challenge	985	native	1,407	multiply	6,230
wing	1,899	delivery	2,821	amount	782	thought	760	Polish	5,821
total	1,039	journalism	4,565	highly	1,324	pineapple	9,278	versus	3,180
singing	6,582	simply	583	item	1,032	cheesecake	14,097	task	1,048
editorial	5,059	everybody	799	pork	4,304	easily	1,213	bravely	14,576
wide	941	bride	4,919	underwear	6,618	secondly	6,993	content	1,741
lighter	8,358	nicely	6,095	guard	1,288	member	285	graduate	2,127

5 | 26 | 19 worse known than freq

2,000-09	2,010-19	2,020-29	2,030-39	2,040-49
donate 4,127	decode 14,250	mentally 4,404	whale 4,267	economically 4,931
chain 1,815	cute 4,383	hunt 5,019	variety 1,172	bean 2,367
return 472	rider 3,854	caller 5,089	coach 1,034	really 143
ideally 7,997	success 777	challenge 985	native 1,407	multiply 6,230
wing 1,899	delivery 2,821	amount 782	thought 760	Polish 5,821
total 1,039	journalism 4,565	highly 1,324	pineapple 9,278	versus 3,180
singing 6,582	simply 583	item 1,032	cheesecake 14,097	task 1,048
editorial 5,059	everybody 799	pork 4,304	easily 1,213	bravely 14,576
wide 941	bride 4,919	underwear 6,618	secondly 6,993	content 1,741
lighter 8,358	nicely 6,095	guard 1,288	member 285	graduate 2,127

15 | 25 | 10 worse known than freq

2,000-09	2,010-19	2,020-29	2,030-39	2,040-49
donate 4,127	decode 14,250	mentally 4,404	whale 4,267	economically 4,931
chain 1,815	cute 4,383	hunt 5,019	variety 1,172	bean 2,367
return 472	rider 3,854	caller 5,089	coach 1,034	really 143
ideally 7,997	success 777	challenge 985	native 1,407	multiply 6,230
wing 1,899	delivery 2,821	amount 782	thought 760	Polish 5,821
total 1,039	journalism 4,565	highly 1,324	pineapple 9,278	versus 3,180
singing 6,582	simply 583	item 1,032	cheesecake 14,097	task 1,048
editorial 5,059	everybody 799	pork 4,304	easily 1,213	bravely 14,576
wide 941	bride 4,919	underwear 6,618	secondly 6,993	content 1,741
lighter 8,358	nicely 6,095	guard 1,288	member 285	graduate 2,127

Spanish Results (4,950-4,999)

4,950-59	4,960-69	4,970-79	4,980-89	4,990-99
attitude 1,156	cane 5,358	hence 3,666	lens 3,570	reinforce 3,351
puzzle 4,291	networking 9,561	logical 4,612	soften 4,916	format 3,758
sole 4,576	unpublished 11,557	separately 5,526	attend 1,169	spine 4,952
tutor 9,285	guardian 5,741	testing 2,517	aim 1,941	cheerleader 10,507
achieve 1,153	metamorphosis 15,000	supply 2,490	sweating 17,550	tough 1,013
contract 1,273	title 5,387	sensationally 42,321	passionately 10,349	suspiciously 12,379
participation 2,228	homophobic 20,584	canyon 5,512	trout 5,098	stab 5,910
feedback 3,918	distribute 2,962	eventually 1,107	predecessor 5,314	persuade 3,286
port 2,734	deadline 4,064	tolerant 9,361	firewall 17,890	seldom 4,926
reply 5,536	bodybuilding 24,019	execute 3,669	tax 509	attorney 1,070

10 | 17 | 23

4,950-59	4,960-69	4,970-79	4,980-89	4,990-99
attitude 1,156	cane 5,358	hence 3,666	lens 3,570	reinforce 3,351
puzzle 4,291	networking 9,561	logical 4,612	soften 4,916	format 3,758
sole 4,576	unpublished 11,557	separately 5,526	attend 1,169	spine 4,952
tutor 9,285	guardian 5,741	testing 2,517	aim 1,941	cheerleader 10,507
achieve 1,153	metamorphosis 15,000	supply 2,490	sweating 17,550	tough 1,013
contract 1,273	title 5,387	sensationally 42,321	passionately 10,349	suspiciously 12,379
participation 2,228	homophobic 20,584	canyon 5,512	trout 5,098	stab 5,910
feedback 3,918	distribute 2,962	eventually 1,107	predecessor 5,314	persuade 3,286
port 2,734	deadline 4,064	tolerance 9,361	firewall 17,890	seldom 4,926
reply 5,536	bodybuilding 24,019	execute 3,669	tax 509	attorney 1,070

LTRC 2020 Tunisia

**Plan is to present one or more lists at
LTRC 2020 in Hammamet**



Vocabulary Challenge

(www.vocabularychallenge.org)



Vocabulary Challenge

Have you ever wondered how many English words you know? Maybe you are a student, you are learning English in your free time, or you use English as part of your job? If so, you can try the British Council's Vocabulary Challenge to check your English word knowledge!

We are looking for Spanish, German and Chinese native speakers to take the Vocabulary Challenge and be a part of this international project.

Take the VOCABULARY CHALLENGE now!

中文

Deutsch

Español