

Putting a test of multiword expressions to a test

Ron Martinez, University of Nottingham

**IATEFL Testing, Evaluation and Assessment SIG and the
University of Innsbruck Conference**

Innsbruck, Austria

16 September 2011

What do you/don't you understand?

Joe suddenly realized Jill was playing the donk, chortlollying the way she was. This was enough to sally the wigs, so he peakled a song and slammed the door behind him.

Weeks later, miggle one by, Joe began to reconsider Jill's frith, and he called her back in though one night after drinking one too many zuppatas.

play the donk (idiom) When a person plays the donk, that person speaks condescendingly.

chortlolly (regular intransitive verb) When a person chortlollies, the lips tighten and the muscles in the cheeks visibly stiffen to try to keep from laughing.

sally the wigs (idiom) Cause one to lose one's temper.

peakle a song (idiom) To leave without saying goodbye.

miggle one by (adverbial expression) Slowly but surely.

frith (countable noun) Intentions, true or imagined.

in though (adverbial expression) Reluctantly.

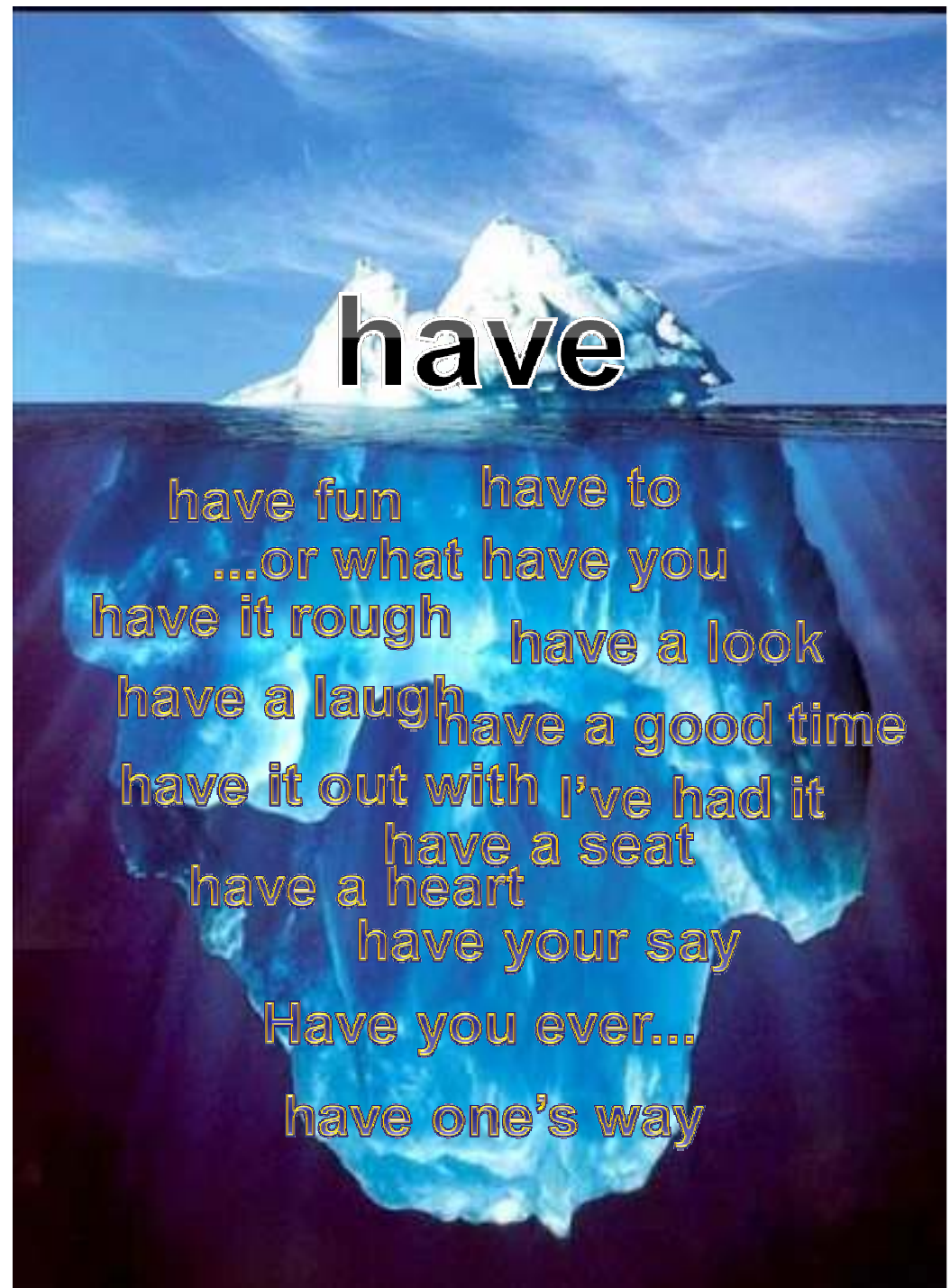
zuppata (noun) A drink composed of a mixture of two or more open and unfinished bottles of wine.



‘frequency’

“There is an obvious payoff for learners of English in concentrating initially on the 2,000 most frequent words, since they have been repeatedly shown to account for at least 80% of the running words in any written or spoken text.” (Read, 2004: 148)

1	THE	6024412
2	BE	3677207
3	OF	3050370
4	AND	2627773
5	I	2615140
6	TO	2601246
7	A	2509255
8	IN	1950387
9	THIS	1705658
10	HE	1174882
11	HAVE	1160957
12	IT	1088473
13	FOR	893866
14	THEY	816758
15	ON	742310
16	YOU	716960
17	WITH	657936
18	AS	650573
19	SHE	640762
20	AT	525832
21	DO	524033



in Organizational

Richard H. Axelrod, Emily Axelrod, Robert

Seventy percent of organizational change efforts fail or fall short of achieving their intended objectives. This figure is even more astounding when you take account of a recent Oxford University study on change, which found that over 68% of employees welcomed meaningful involvement in change. Clearly, we are missing the boat when it comes to effective change, and our clients are not getting what they paid for.

We believe that there is a relationship between the failure rate of

The Fe

As the
velopin
way ou
emerge

• • • • •

Heed organizatio

RON MARTINEZ

Lexical Profile using *VocabProfile*

Frequency Band	Words (types)	Text coverage (tokens)
0-1000	<i>a account and are boat change clearly comes effective efforts employees even fail fall figure for found getting in is it meaningful missing more not number of on or our over paid recent seventy short study take that the their they this to university we welcomed what when which you</i>	86.57%
1001-2000	<i>intended</i>	1.49%
Academic Word List	<i>achieving involvement percent</i>	4.48%
Off list	<i>astounding clients objectives organizational Oxford</i>	7.46%
Words in Top 2,000:		88.06%
+ AWL words:		4.48%
Total text coverage:		92.54%

in Organizational

Richard H. Axelrod, Emily Axelrod, Robert

Seventy percent of organizational change efforts fail or fall short of achieving their intended objectives. This figure is even more astounding when you take account of a recent Oxford University study on change, which found that over 68% of employees welcomed meaningful involvement in change. Clearly, we are missing the boat when it comes to effective change, and our clients are not getting what they paid for.

We believe that there is a relationship between the failure rate of

The Fe

As the
velopin
way ou
emerge

• • • • •

Heed organizatio

Ron Martinez

Lexical Profile using *VocabProfile*

Frequency Band	Words (types)	Text coverage (tokens)
0-1000	<i>a account and are boat change clearly comes effective efforts employees even fail fall figure for found getting in is it meaningful missing more not number of on or our over paid recent seventy short study take that the their they this to university we welcomed what when which you</i>	86.57%
1001-2000	<i>intended</i>	1.49%
Academic Word List	<i>achieving involvement percent</i>	4.48%
Off list	<i>astounding clients objectives organizational Oxford</i>	7.46%
Words in Top 2,000:		88.06%
+ AWL words:		4.48%
Total text coverage:		92.54%

Multiword-Inclusive Profile

Frequency Band	Words and Phrases (types)	Text coverage (tokens)
0-1000	<i>a and are change clearly effective efforts employees even fail figure for found getting in is meaningful more not number of on or our over paid recent seventy study that their they this university we welcomed what when which you</i>	67.16%
1001-2000	<i>intended</i>	1.49%
Academic Word List	<i>achieving involvement percent</i>	4.48%
Off list	<i>astounding clients fall short of missing the boat objectives organizational Oxford take account of when it comes to</i>	26.87%
Words in Top 2,000:		68.65%
+ AWL words:		4.48%
Total text coverage:		73.13%

Ron Martinez



‘frequency’: potentially problematic
from both the perspective of learner
and teacher/tester.

Martinez and Murphy (2011)

- 101 adult Brazilian learners of English ('intermediate' or higher).
- Within-groups, paired samples of reading comprehension measures on a two-part reading test.
- All texts on both test parts written 'symmetrically', using exact same pool of top 2,000 word families in English (BNC).

Ron

Let me tell you about my home. It's on this little hill out in the country. But I'm not far from the city (I don't like the city – do you?) – not much time to get here. I can't wait to show you a photo... or you can call me to come over to see in person! 07786 237 679

exact
same
words

all very
frequent
words
(top
2,000)

I don't get out much – it's about time I do. I'm not from here – this country or city. (But I like this country.) I'm far from home. I'm a little over the hill, let me tell you, but you can't tell! (I can show you my photo, or wait to come see me in person!) Call me on 07786 554 0978

Test Overview

- Part 1: 4 texts, 7 questions each – compositional formulations (meanings transparent from individual words).
- Part 2: 4 texts, 7 questions each, exact same words – less compositional.
- Rating scale for self-reported comprehension after each text.

I don't get out much – it's about time I do. I'm not from here – this country or city. (But I like this country.) I'm far from home. I'm a little over the hill, let me tell you, but you can't tell! (I can show you my photo, or wait to come see me in person!) Call me on 07786 554 0978

1. ☐ He wants to go out but has a problem with time.
2. ☐ He is foreign.
3. ☐ He lives in a remote area.
4. ☐ He wants to keep his location a secret.
5. ☐ He thinks he looks younger than his age.
6. ☐ He probably lives in an area with hills.
7. ☐ He lives on the hill, but not on top of it.

My comprehension of this text: 5% 25% 50% 75% 100%

The results

	Min.	Max.	Mean	SD
Part 1 Total	18	28	24.09	2.44
Part 2 Total	6	25	14.76	3.93

$t = 24.10$ ($p \leq 0.001$), eta squared = 0.828

Reported Comprehension vs. Actual Comprehension

- No statistically significant difference for Part 1 (87.38% reported vs 86.03% actual).
- Reported comprehension significantly overestimated in Part 2 ($t = 3.95$, $p \leq 0.001$, eta squared = 0.07) – **60.29%** reported vs **52.58%** actual.

I don't like to go out much. The thing is, I have two left feet and I need to watch my spending. Now, for what it's worth, every so often I do have a ball playing football with my kids (my kids see me through the day) and even having one drink on occasion. (And not two at a time!) But I never get carried away or lose it. Having a good time doesn't have to be hard or cost you an arm and a leg. (Mind you, I'm not flat broke!) I'm a real team player, looking out for others all the time – especially my special kids. With my kids, a good time is to be had every time.

- 15. ☐ She does not know how to dance well.
- 16. ☐ She thinks football and drinks are worth the money.
- 17. ☒ She frequently plays football with her children.
- 18. ☐ Her children are with her most of the day.
- 19. ☒ She drinks when it is a special occasion.
- 20. ☐ Her arms and her legs have been injured, but not completely broken.
- 21. ☐ She is looking for more players for the football team.

My comprehension of this text: 5% 25% 50% 75% 100%

‘on occasion’

		T2_19			
		Frequency	Percent	Valid Percent	Cumulative Percent
INTERMEDIATE	Valid Incorrect	22	46.8	46.8	46.8
	Correct	25	53.2	53.2	100.0
	Total	47	100.0	100.0	

		T2_19			
		Frequency	Percent	Valid Percent	Cumulative Percent
HIGHER	Valid Incorrect	25	46.3	46.3	46.3
	Correct	29	53.7	53.7	100.0
	Total	54	100.0	100.0	

The Yes-No Test (Meara, 1992)

1 ☐ dog

4 ☐ steam

7 ☐ leaf

10 ☐ tap

13 ☐ albuolic

2 ☐ arm

5 ☐ invent

8 ☐ damage

11 ☐ pig

14 ☐ truth

3 ☐ powling

6 ☐ empty

9 ☐ scissors

12 ☐ real

15 ☐ boil

The Vocabulary Levels Test (Nation, 1983; Schmitt, Schmitt & Clapham, 2001)

1. original

2. private

3. royal

4. slow

5. sorry

6. total

_____ first

_____ not public

_____ all added together

Vocabulary Size Test (Nation & Beglar, 2007)

First 1000

1.	see: They saw it.	
	a.	cut
	b.	waited for
	c.	looked at
	d.	started

2.	time: They have a lot of time .	
	a.	money
	b.	food
	c.	hours
	d.	friends

Research question

How can a test be devised that assesses knowledge of multiword expressions in the same or similar way as current widely-used vocabulary tests?

Challenges

1. Narrowing down the phraseological field (i.e. which formulaic sequence?)
2. Pinning down the extent (i.e. where do you stop?)
3. Finding the expressions (i.e. what tools and resources can be used?)
4. Adopting an appropriate test format (i.e. how to test the sequences?)

Challenges

1. Narrowing down the phraseological field (i.e. which formulaic sequence?)
2. Pinning down the extent (i.e. where do you stop?)
3. Finding the expressions (i.e. what tools and resources can be used?)
4. Adopting an appropriate test format (i.e. how to test the sequences?)

The Yes-No Test (Meara, 1992)

1 ☐ dog

4 ☐ steam

7 ☐ leaf

10 ☐ tap

13 ☐ albuolic

2 ☐ arm

5 ☐ invent

8 ☐ damage

11 ☐ pig

14 ☐ truth

3 ☐ powling

6 ☐ empty

9 ☐ scissors

12 ☐ real

15 ☐ boil

The Vocabulary Levels Test (Nation, 1983; Schmitt, Schmitt & Clapham, 2001)

1. original

2. private

3. royal

4. slow

5. sorry

6. total

_____ first

_____ not public

_____ all added together

Vocabulary Size Test (Nation & Beglar, 2007)

First 1000

1.	see: They saw it.	
	a.	cut
	b.	waited for
	c.	looked at
	d.	started

2.	time: They have a lot of time .	
	a.	money
	b.	food
	c.	hours
	d.	friends

More compositional?

Less compositional?



at all times

at all costs

at all

Meaning still retained when each
lexical word replaced with its own
definition (Grant & Bauer, 2004)

A 'phrasal expression'

- A fixed or semi-fixed sequence of two or more co-occurring but not necessarily contiguous words with a cohesive meaning or function that is not easily discernible by decoding the individual words alone.
- *take place, to a large extent, take sth over*

Challenges

1. Narrowing down the phraseological field (i.e. which formulaic sequence?)
2. Pinning down the extent (i.e. where do you stop?)
3. Finding the expressions (i.e. what tools and resources can be used?)
4. Adopting an appropriate test format (i.e. how to test the sequences?)

Frequency

- VLT stopped at 5000 word frequency band “represents the upper limit of general high-frequency vocabulary” (Read, 2000: 119)
- a vocabulary size of 5000 allows for “pleasurable reading” of simple fiction (Hirsh & Nation, 1992)
- the English Profile Wordlist project has 4667 entries through B2 (CEFR)
- by advanced levels, students “would probably be expected to recognize over 4500” word families (Milton, 2009: 180)

1	THE	6024412
2	BE	3677207
3	OF	3050370
4	AND	2627773
5	I	2615140
6	TO	2601246
7	A	2509255
8	IN	1950387
9	THIS	1705658
10	HE	1174882
11	HAVE	1160957
12	IT	1088473
13	FOR	893866
14	THEY	816758
15	ON	742310
16	YOU	716960
17	WITH	657936
18	AS	650573
19	SHE	640762
20	AT	525822

BNC Band Cut-off Points

Frequency band	Token frequency cut-off	Frequency band	Token frequency cut-off
----------------	-------------------------	----------------	-------------------------

1,000	12,639 +	8,000	434 +
2,000	4,491 +	9,000	356 +
3,000	2,089 +	10,000	295 +
4,000	1,210 +	11,000	249 +
5,000	787 +	12,000	213 +
6,000	620 +	13,000	184 +
7,000	547 +	14,000	162 +

1	THE	6024412
2	BE	3677207
3	OF	3050370
4	AND	2627773
5	I	2615140
6	TO	2601246
7	A	2509255
8	IN	1950387
9	THIS	1705658
10	HE	1174882
11	HAVE	1160957
12	IT	1088473
13	FOR	893866
14	THEY	816758
15	ON	742310
16	YOU	716960
17	WITH	657936
18	AS	650573
19	SHE	640762
20	AT	525822

BNCindex_index_2-4-word clusters.lst

File Edit View Compute Settings Window Help

N	Word	Freq.	%	Texts	% emmas	Set
1	OF THE	763,518	0.77	4,049	100.00	
2	IN THE	517,267	0.52	3,970	98.05	
3	TO THE	289,777	0.29	3,915	96.69	
4	ON THE	220,794	0.22	3,894	96.17	
5	AND THE	196,164	0.20	3,870	95.58	
6	TO BE	188,925	0.19	3,866	95.48	
7	FOR THE	166,668	0.17	3,853	95.16	
8	AT THE	150,944	0.15	3,867	95.51	
9	IT IS	137,121	0.14	3,773	93.18	
10	BY THE	130,745	0.13	3,700	91.38	
11	THAT THE	129,039	0.13	3,672	90.69	
12	WITH THE	128,384	0.13	3,817	94.27	
13	IT WAS	127,276	0.13	3,657	90.32	
14	OF A	124,990	0.13	3,756	92.76	
15	FROM THE	124,636	0.13	3,787	93.53	
16	IN A	114,398	0.12	3,794	93.70	
17	AS A	88,236	0.09	3,662	90.44	
18	IS A	78,247	0.08	3,688	91.08	
19	WITH A	78,166	0.08	3,681	90.91	
20	FOR A	72,791	0.07	3,728	92.07	
21	HE WAS	72,233	0.07	2,883	71.20	
22	HAVE BEEN	70,026	0.07	3,620	89.40	
23	WILL BE	69,603	0.07	3,445	85.08	
24	WAS A	65,113	0.07	3,387	83.65	
25	HAD BEEN	64,049	0.06	2,960	73.10	
26	IS THE	63,326	0.06	3,637	89.82	
27	ONE OF	63,092	0.06	3,653	90.22	
28	AS THE	62,618	0.06	3,484	86.05	
29	THE FIRST	61,412	0.06	3,635	89.78	
30	TO A	59,227	0.06	3,607	89.08	
31	THE SAME	57,981	0.06	3,630	89.65	
32	HAS BEEN	57,454	0.06	3,327	82.17	
33	AND A	57,347	0.06	3,644	90.00	

frequency alphabetical statistics filenames notes

4,258,784 Type-in

BNCindex_index_2-4-word clusters.lst

File Edit View Compute Settings Window Help

N	Word	Freq.
1	OF THE	763,518
2	IN THE	517,267
3	TO THE	289,777
4	ON THE	220,794
5	AND THE	196,164
6	TO BE	188,925
7	FOR THE	166,668
8	AT THE	150,944
9	IT IS	137,121
10	BY THE	130,745
11	THAT THE	129,039
12	WITH THE	128,384
13	IT WAS	127,276
14	OF A	124,990
15	FROM THE	124,636

Initial data deletion using criteria

632	ARE IN	7,953	2,422	59.82
633	AND AN	7,942	2,452	60.56
634	AS A RESULT	7,939	1,990	49.15
635	SHE WOULD	7,939	1,206	29.79
636	BUT SHE	7,930	1,260	31.12
637	RIGHT TO	7,893	2,020	49.89
638	A COUPLE	7,863	1,861	45.96
639	WHO IS	7,863	2,131	52.63
640	TO MEET	7,862	2,249	55.54
641	THE THIRD	7,848	2,219	54.80
642	EXPECTED TO	7,840	2,029	50.11
643	LOOKED AT	7,839	1,544	38.13
644	THE SOUTH	7,826	1,662	41.05
645	IN ADDITION	7,822	1,813	44.78
646	THE ROYAL	7,817	1,481	36.58

7

June 14 General situation: Rain **at first**, some bright spells. Most of

8

~~Walter Love have an identity problem? At first sight it would seem he has, for~~

9

laser before their eyes. He says that **at first** clients refused to believe it could

10

the grim line of his mouth but his eyes **at first** showed open admiration as they

11

been total and, she had believed, final. **At first** the agony had been unbearable.

12

~~these key sequences to the CLI. At first glance, the following program line~~

13

so unexpectedly I probably acted dumb **at first**. He stood with his hands and his

14

had gone to him, the Prince had in fact **at first** turned the suggestion down. It

15

stayed like that for the rest of the day. **At first** it was cool and rather pleasant

16

~~something put up there, beyond your at first. Yes,, James. Erm, James put a~~

17

in the past,(though this may be difficult **at first**), learning to ask "when" and "for

18

felt in a daze. She had been convinced **at first** that it had all been a mistake,

19

for Bunny's truck. He looked suspicious **at first**, and so did the Japanese waiters

20

the road between the parked cars and **at first** they couldn't see a space

21

a ghastly scream from inside the room. **At first** Robert thought someone must

PHRASE List sample

3362	IN ADVANCE	1983	It's something you should do in advance .
3379	ON THE PART OF	1965	There are no barriers on the part of the government.
3387	BRING UP	1958	Did you manage to bring up the holiday pay?
3388	TAKE OFF	1957	I was hungry by the time we took off .
3390	SO AS TO	1954	He cluttered his desk so as to appear busy.
3396	TAKE ADVANTAGE	1949	Some people just want to take advantage .
3399	SHORT OF	1944	Short of calling the doctor, I don't know what to do.
3402	OVER THE YEARS	1942	She has over the years visited many cities.
3406	SWITCH ON	1935	It's plugged in but it won't switch on .
3417	BY NO MEANS	1925	We have by no means concluded the matter.
3431	COULD HARDLY	1911	I could hardly wait for the weekend.
3446	COME UP WITH	1898	Is that the best you could come up with ?
3448	IN QUESTION	1898	The book in question was published last year.
3450	IN THE FIRST PLACE	1897	He couldn't see how it got there in the first place .
3456	PROVE TO BE	1893	The bicycle proved to be of immense value.
3457	IN COMMON	1890	What they have in common is their stamina.
3461	NO MATTER	1888	No matter what they say, it's down to you now.

single word – multiword expression frequency matching

wordlist

4719.	CULT	1061
4720.	DESCENT	1061
4721.	STOCKING	1061
4722.	BELLY	1060
4723.	NUTRITION	1060
4724.	BRACKET	1059
4725.	SOFA	1059

integrated

4719.	CULT	1061
4720.	DESCENT	1061
4721.	FOR GOOD ('FOREVER')	1061
4722.	STOCKING	1061
4723.	BELLY	1060
4724.	NUTRITION	1060
4725.	BRACKET	1059
4726.	(BE) OPPOSED TO	1059
4727.	SOFA	1059

Challenges

1. Narrowing down the phraseological field (i.e. which formulaic sequence?)
2. Pinning down the extent (i.e. where do you stop?)
3. Finding the expressions (i.e. what tools and resources can be used?)
4. Adopting an appropriate test format (i.e. how to test the sequences?)

Pilot 1 (n=10): VLT format

From 1,000 to 2,000

- 1 as to
- 2 a couple of _____ until now
- 3 as a result _____ because of that
- 4 for instance _____ more than one
- 5 long term
- 6 so far

From 2,000-3,000

- 1 turn out
 - 2 work out _____ indicate your answer is no
 - 3 carry on _____ try to see or inspect something
 - 4 shake your head _____ continue
 - 5 have a look
 - 6 look forward to
-

1 happen to	
2 make sense	_____ by chance
3 amount to	_____ find
4 get on with	_____ become
5 come across	
6 bear in mind	

1 a variety of

2 or so

3 in time

4 whether or not

5 in a way

6 above all

_____ if

_____ many

_____ maybe a little more or less

Vocabulary Size Test (VST) (Nation & Beglar, 2007)

First 1000

1.	see: They saw it.	
	a.	cut
	b.	waited for
	c.	looked at
	d.	started

2.	time: They have a lot of time .	
	a.	money
	b.	food
	c.	hours
	d.	friends

Pilot 2: VST + VLT (n=34)

4.	to do with: It is to do with money.	
	a.	making
	b.	for
	c.	about
	d.	our

Pilot 2 (VST-VLT comparison)

- 48 overlapping items, counterbalanced forms (VLT/VST)
- immediate post-test interviews
- VST format 100% preferred by candidates

declared knowledge discrepancies

	VST	VLT
Discrepancies	11	77
(max.=48)	M = 1.50	M = 8.80

- Vocabulary Levels Test (VLT) version significantly more prone to knowledge discrepancies ($t = 5.439$, $p \leq 0.001$)

Field test (n = 2203)

Test Version	N	Mean	SD
A	742	22.67	5.30
B	731	22.32	5.76
C	730	21.95	5.59

Freq.	Versio n A	M	SD	Versio n B	M	SD	Versio n C	M	SD
1K		5.50	0.87		4.78	1.26		4.25	0.97
2K		5.05	1.20		5.17	1.14		4.65	1.41
3K		4.33	1.34		4.63	1.44		4.72	1.59
4K		4.21	1.65		3.52	1.62		4.01	1.56
5K		3.60	1.65		4.22	1.67		2.32	1.63

12. at once: I did it **at once**.

- a. one time
- b. many times
- c. early
- d. immediately

☐☐☐☐

K3

12.	at once: I did it at once .		Facility	Upper	Lower	D
	a.	one time	.47	.16	.78	-.62
	b.	many times	.00	.00	.00	.00
	c.	early	.02	.00	.06	-.06
	d.	immediately	.43	.81	.16	.65
		No attempt		4 (2%)	29 (16%)	

K2

3	so far: It's good so far.		Facility	Upper	Lower	D
	a.	until now	.90	1.00	.75	.25
	b.	but not really	.04	.00	.08	-.08
	c.	sometimes	.01	.00	.02	-.02
	d.	from a distance	.05	.00	.15	-.15
		No attempt		0 (0%)	12(5%)	

K1

14	used to: I used to go.		Facility	Upper	Lower	D
	a.	want to	.12	.01	.29	-.28
	b.	did before	.26	.55	.07	.48
	c.	usually	.56	.40	.54	-.14
	d.	always	.07	.05	.09	-.04
		No attempt		1 (0.5%)	14 (7.2%)	

	Answer type totals	Combined totals*
Answer type (consistent)		
‘0’ = Incorrect answer and translation	33	740 (consistent)
‘1’ = Correct answer and translation	707	
Answer type (discrepant)		
‘2’ = Incorrect answer, correct translation	6	8 (discrepant)
‘3’ = Correct answer, incorrect translation	2	

Summary

- Vocabulary in testing – both discrete and embedded – can and should be inclusive of multiword expressions.
- One early version of the Phrasal Vocabulary Size Test has gone through **one** early phase of validation.
- Test does appear to assess knowledge of phrasal expressions on PHRASE List and could be used in conjunction with the VST, for example.
- But so what? Still need to understand what knowledge of phrasal expressions means in the bigger SLA picture.



99^p

CHEESEBURGER

THAT'S McDONALD'S... AND THEN SOME.™

Served after 10:30am. Products on sale may be restricted at different times. Price and participation may vary.
© 2009 McDonald's.

aexrm3@nottingham.ac.uk

<http://nottingham.academia.edu/RonMartinez>

Sample score sheet

PVST Level	No. correct		Estimated knowledge of PHRASE List expressions	Total possible
First 1000	9		28	32
Second 1000	8		67	84
Third 1000	5		65	129
Fourth 1000	2		31	157
Fifth 1000	1		10	103
		Estimated total	201	505

Test	N	Participant Minimum	Participant Maximum	Mean	SD
PVLT Discrepancies	10	3	15	8.80	4.18
PVST Discrepancies	10	0	3	1.50	0.97

Frequency Band	PVLT	PVST
1K	3	4
2K	21	0
3K	19	2
4K	21	2
5K	13	3
Total (Max.= 480)	77	11

Ron
Martinez

Summary of Test 1 and Test 2 text data*

	Total word count	Total clause count	Top 1,000 words	Top 1,001 – 2,000
Test 1	416	56	95.7373%	2.7650%
Test 2	412	54	95.7373%	2.7650%

* Measured using Vocabulary Profiler

PHRASE List grammatical analysis

Band	NP	VP	Adv.	Adj.	DET./PRO	Other
1K (k = 32)	0	7	23	0	1	1
2K (k = 85)	1	38	26	1	14	5
3K (k = 128)	2	45	63	3	12	3
4K (k = 158)	4	38	97	3	12	4
5K (k = 102)	4	32	56	2	5	3
Total = 505	11	160	265	9	44	16
Cum. %	2.17%	31.68%	52.47%	1.78%	8.71%	3.16%