Language Testing Forum 2015

Assessing vocabulary: a multidimensional framework combining corpus data and teacher judgements

Veronica Benigno & John de Jong (Pearson)

21 November, 2015
Department of Education, University of Oxford

Goals of our talk

- Argue for the need of a **valid vocabulary framework** which complements the functional guidance found in the CEFR (Council of Europe, 2001) descriptors
- Discuss key **evidence and limitations of current research** on vocabulary assessment
- Present a **meaning-based** and **CEFR-linked vocabulary database** to inform teaching, course design, and assessment

Overview of current research

- **Positive relationship between vocabulary knowledge & reading** (Anderson and Freebody, 1979/1981; Alderson, 2000; Quian, 2002; Perfetti, 2007)

- **Vocabulary as a multidimensional construct: size + depth** (Read, 2004; Meara, 2005; Daller et al., 2007; Schmitt et al., 2010)

- **There is some (conflicting?) evidence about size requirements**
  - 5-7k spoken & 8-9k written word families for non-assisted comprehension (Nation, 2006)
  - 10k lemmas for academic studies (Hazenberg & Hulstijn, 1996)

- **And concerns exist about the unit of counting** (Bogaards, 2001)
  - Word-families? *NATION. NATIONAL? NATIONWIDE? NATIONALISM?*
  - Lemmas? *BALL*. The round object? The dance event? The tip of the finger?

Some limitations

- **Frequency as main principle to set vocabulary requirements**
  - Cost-benefit argument: first 1k words - 84% coverage of informal spoken (Laufer & Nation, 1999)
  - Psycholinguistic argument: input frequency (Paradis, 2007; Ellis, 2003)

  <u>BUT it's not only a matter of frequency!</u>

- **Functional aspects of vocabulary learning often neglected**
  - Vocabulary usefulness/coreness (Carter, 1987; Stubbs, 2009, Wray, 2002)
  - Incremental learning of word meanings (Brent, 2009; Henriksen, 1999; Nagy and Scott, 2000)

What do we need then?

- **A unitary model of lexical proficiency which** takes into account polysemy, collocations, chunks, 'contextual knowledge' (Perfetti et. al., 2010)

- A **meaning-based approach to vocabulary:** what word meanings should learners master at different proficiency levels?

"An assessment of the number of meanings a reader knows enables a remarkably accurate prediction of this individual's ability to comprehend discourse" (Anderson and Freebody,1979)

"[…] the concept of scalar word difficulty and current corpus-based methods for its estimation are inadequate. We propose a new paradigm called word meaning maturity which tracks the degree of knowledge of each word at different stages of language learning" (Kireyev & Landauer, 2011)

A CEFR-linked vocabulary framework

A graded lexical inventory of **general English for adult learners**

- **Aligned to the CEFR and the Global Scale of English**
- **Combines frequency + communicative usefulness** – as per judgments provided by pool of teachers
- **Total of 36,000 word-meanings** (28,000 lemmas)
- **Searchable online** by keyword, part of speech, topic/subtopic, collocations, sentence examples, proficiency level, etc.
- **To help learners, teachers, and test/materials designers** identify level-appropriate vocabulary

Methodology

- STEP 1 – Corpus analysis
- STEP 2 – Semantic annotation
- STEP 3 – Teacher ratings
- STEP 4 – Vocabulary scaling

Step 1– Corpus analysis

- **Choice of corpus (L1 data; 2.5 billion words; SP + WR)**
  - **LCN** (Longman Corpus Network). A balanced and representative corpus of British/American English
  - **UKWaC** (Baroni et al., 2009). A web-based corpus crawled from the .uk domain, very large in size
  - **COCA** (Corpus of Contemporary American English). Spoken component selected for this study

- **Data processing and creation of frequency list (28k lemmas)**
  - Top 10k items in the corpus + 25 k items in LASDE (*Longman Active Study Dictionary of English*)

STEP 2– Semantic annotation

- Structure vocabulary around pedagogically relevant sets using the **CoE Vantage categorization** in:
  - **Specific Notions** (Topics)

    E.g. *Fork* > FOOD&DRINKS_tableware, SPORT&HOBBIES_gardening, TRAVEL_directions
  - **General Notions**

    E.g. *Somewhere* > SPATIAL_location&position
  - **Functions**

    E.g. *I am sorry* > INTERACTING_making an apology

- **Manual tagging of about 36k word meanings in DPS** by team of lexicographers

14 Specific Notions – Vantage (Council of Europe, 2001)

"[…] those which deal with more concrete details of the here-and-now. They are likely to occur in relation to certain transactions and certain topics"

Appendix A Lexical exponents of specific notions for *Vantage*

10 Food and drink

10.1 types of food and drink (See also General Notions 5.1.6.)

food, meal, snack
to eat, to drink
names of meals e.g. breakfast, lunch, tea, dinner, supper
to have breakfast, etc.
course (as in: What's the main course?)
vegetables
names of vegetables e.g. beans, cabbage, carrots, cauliflower, mushrooms, peas, potatoes, spinach
chips
banana, grape, lemon, nut, orange, raisin, pear, strawberry, raspberry
cereals
crisps
ice-cubes
spice(s), to spice
names of spices e.g. curry, garlic, mustard, pepper, salt, oil, vinegar
flavour
spaghetti, macaroni
flour
bread
slice, piece (as in: Would you like a piece of cake?)
bread roll
toast, sandwich
bun

PEARSON

8 General Notions - Vantage (Council of Europe, 2001)

" […] those which a learner may need to handle whatever the situational context in which a communication act may take place and are largely concerned with abstract, relational concepts"

General notions for *Vantage* including recommended exponents

2 Spatial

2.1 *location*
noun: location, place, position, site, situation
adverbs: here, there, here and there, everywhere, somewhere, nowhere, (not) anywhere, where?, wherever; inside, outside; in the east/north/south/west
verbs: to be (as in: Harrods is in Knightsbridge.), to be situated, to lie
determiners: this, that, these, those

2.2 *relative position*
prepositions: above, against, among, at, at the end of, at the

[illegible] enter, to drop, to fall, to flow, to get up, to go, to hurry, to leave (as in: We have to leave now.), to lie down, to move (as in: The car did not move.), to pass (as in: You pass the railway station.), to quit, to remove, to roll, to run, to stand still, to start, to step, to stop, to walk

2.5 *direction*
direction (as in: In which direction is Slough?), destination
direct (as in: Is there a direct route to London from here?)
adverbs: away, back, backward(s), (anti-)clockwise,

PEARSON

6 Functions - Vantage (Council of Europe, 2001)

" […] denoting in a general way what people do by means of language"

Language functions for *Vantage* and recommended exponents with examples

- Is anything wrong?
- What's the matter/problem?
- What's wrong?

2.5.24 complaining
- I have a complaint (to make).
- I'm sorry, but . . .
- I don't want to complain, but . . .
  Then expressions and structures as per 2.5.22

2.5.25 expressing bad temper
- I'm in a bad mood/temper (just now/today).

trust).

2.5.28 expressing interest
- Really!
- Fascinating!
- Is that so!
- Well, well.
- (How) (very) interesting!
- I am (very) interested in NP/VPgerund
  I am interested in old stamps.
- NP interest(s) me (greatly)
  Greek men interest me greatly.

STEP 3 – Teacher ratings

- **A random 10 out of 19 teachers rated usefulness of vocabulary by meaning** :

  "plate" (1): *a flat dish that you use for eating or serving food*
  "plate" (2): *in baseball, the place where the person hitting the ball stands*

- […] establish the **relative importance of vocabulary items in language production** to successfully convey a particular message.

1 = Essential	*words learners would want to acquire first*
2 = Important	*words that become necessary at a next stage*
3 = Useful	*words enabling more detailed and specific language*
4 = Nice to have	*words to express concepts more accurately*
5 = Extra	*words some language users will use occasionally*
99 "Escape"	*words which are impossible to rate - you have never heard of the word before or you cannot decide between widely different ratings*

STEP 4 – Scaling

- **Combine frequency data and teacher ratings to produce a weighted algorithm to grade vocabulary**

    1. Statistical analysis and data cleaning
    2. Theoretical assumptions based on research evidence
    3. Data modelling

1. Statistical analysis

The first action is to check the soundness of the rating data

Measures per rater

- overall average of a rater compared with other raters
- standard deviation of the raters' ratings
- correlation of a rater's ratings with average of all raters
    - One rater removed

Measures per word: **450,157** ratings gathered

- remove rating if >1.5 distance from average rating
    - 4.7% ratings removed
- Percentage of ratings within two adjacent categories
    - words <70% in two adjacent categories: 1575 (=4.9%)
- Correlation of ratings and frequency data
    - r= 0.44

Combine ratings and frequency data

- Frequency rank data ranging from 3 to 54,496
- Rescale to 1-5 scale
- Then combine average ratings with Freq Rank
- Weight rating data for their reliability

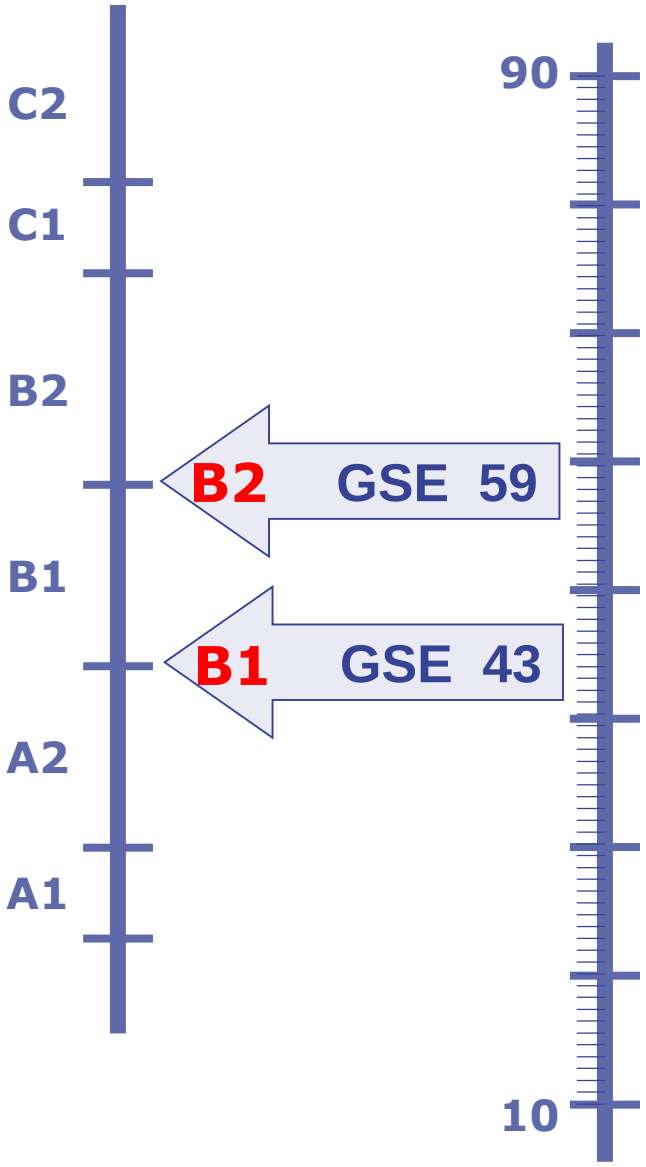
Rating average = Ra

Reliability of rating data = r_{Rating}

Freq Rank= F_{rank}

$$Ra \times r_{Rating} + F_{rank\ m} \times (1 - r_{Rating})$$

CEF Global Scale of English



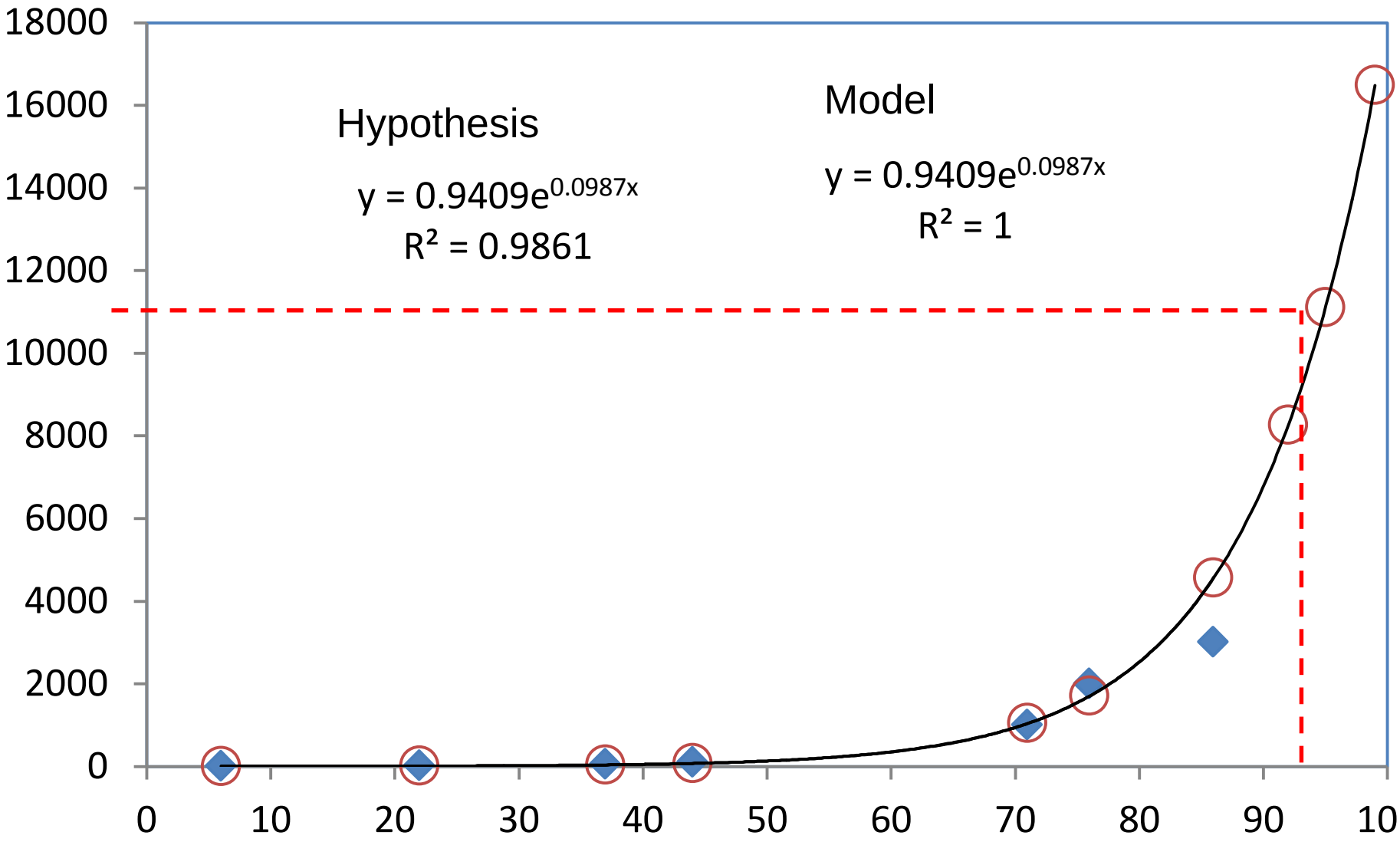
PEARSON

Grabe, 2008 (Chapter 13)

Table 13.2 *Word frequency coverage of academic texts (Nation, 2001, 2004; Schmitt, 2000)*

the	6–7% of total word coverage
top 10 words	22% of coverage
top 50 words	37% of coverage
top 100 words	44% of coverage
top 1,000 words families	71% of coverage
top 2,000 words families	76% of coverage
BNC 3000 word families	86% of coverage

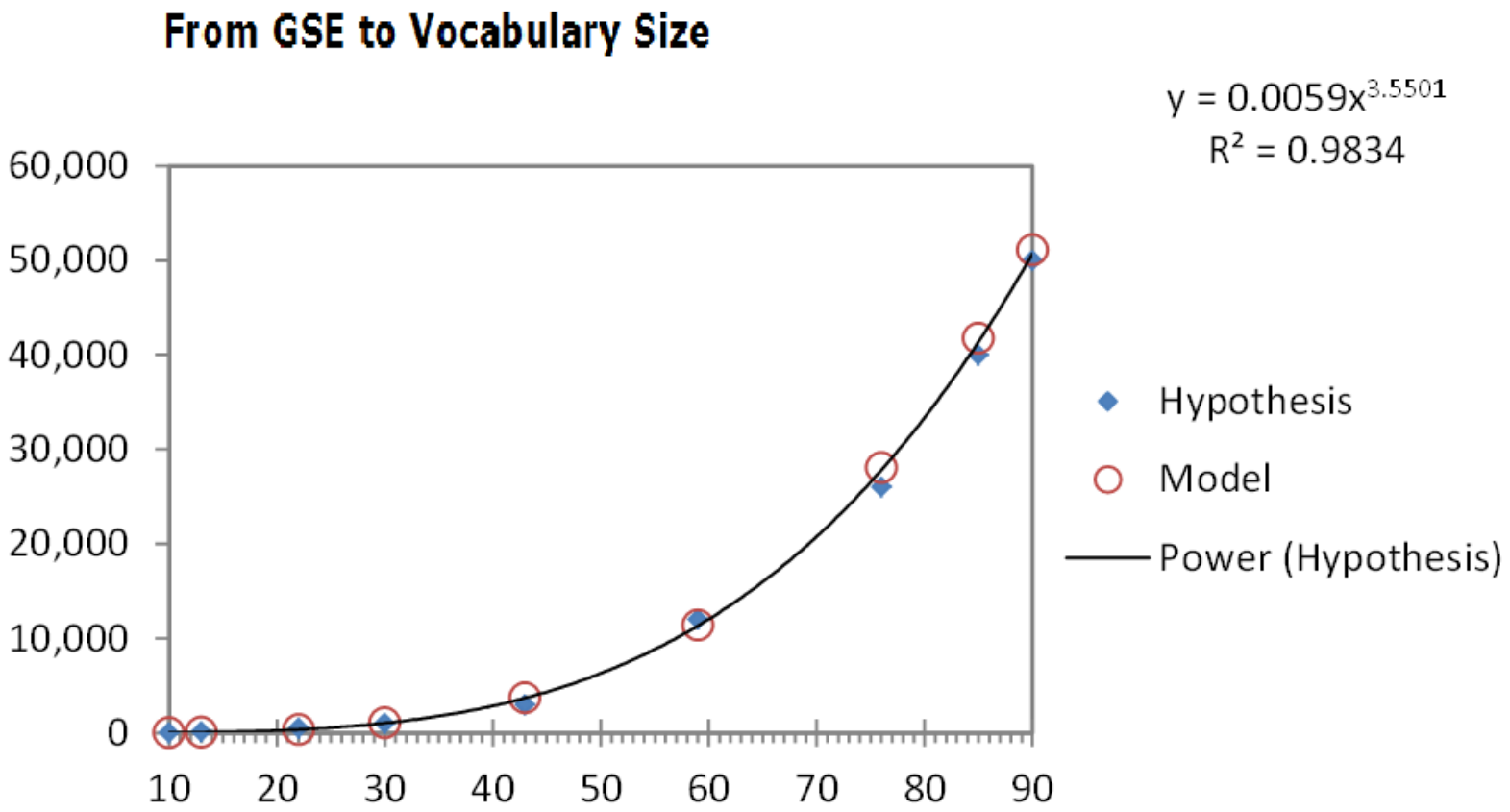
PEARSON


Hypothesis
y = 0.9409e^0.0987x
R² = 0.9861
Model
y = 0.9409e^0.0987x
R² = 1
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
10
20
30
40
50
60
70
80
90
100

2. Theoretical assumptions / Hypothesis

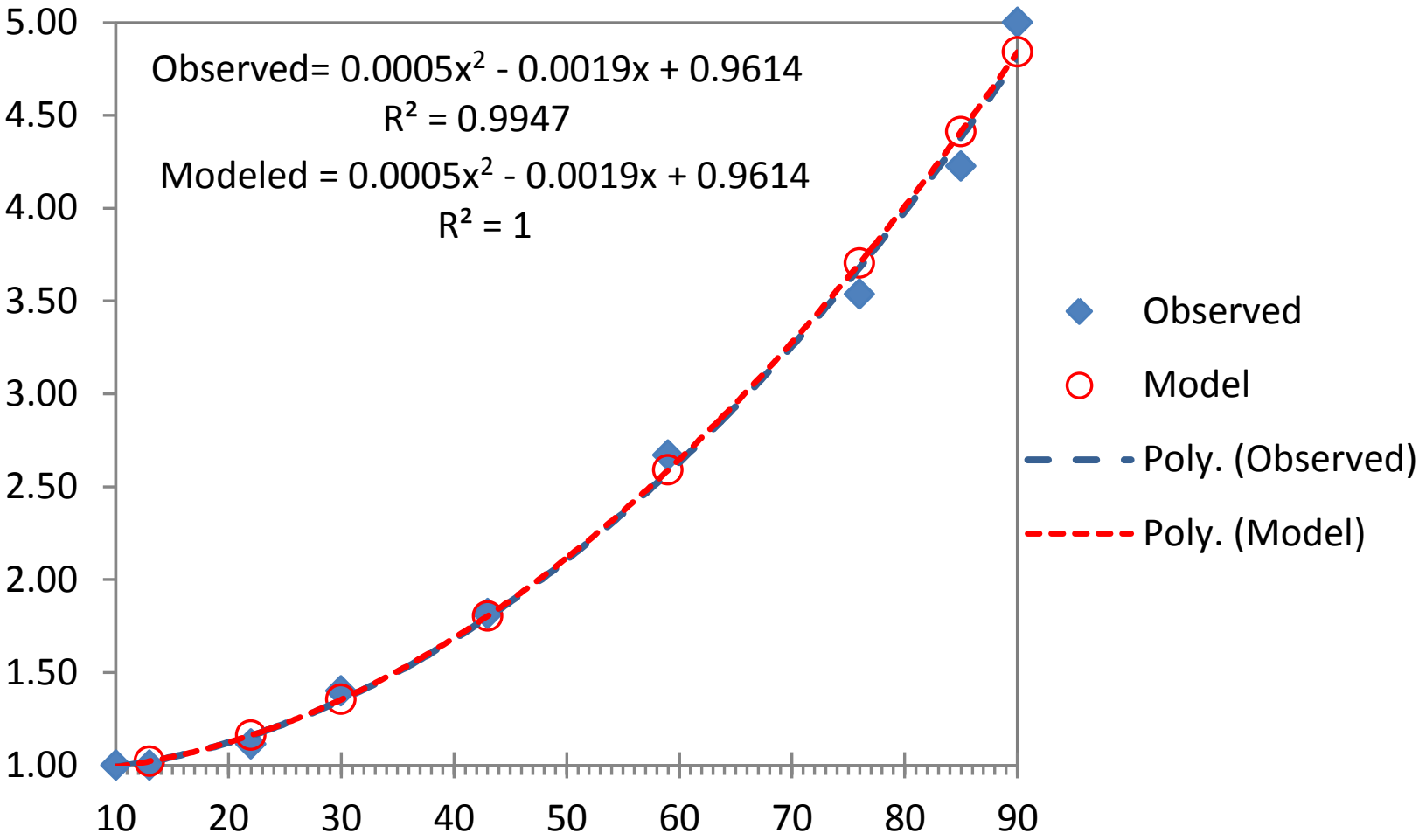
Total number of vocabulary entries				31651		
CEFR	GSE	Hypothesis	Model	NewWords	Proportion	N entries
Start	10	10	21	21	0.0004	13
Tourist	13	100	53	32	0.0006	20
A1	22	500	344	291	0.0057	180
A2	30	1,000	1,035	691	0.0135	428
B1	43	3,000	3,714	2,679	0.0524	1659
B2	59	12,000	11,417	7,703	0.1507	4769
C1	76	26,000	28,050	16,633	0.3254	10298
C2	85	40,000	41,733	13,684	0.2677	8472
Finish	90	50,000	51,122	9,389	0.1837	5813
Total				51,122	1.0000	31,651

3. Data modelling

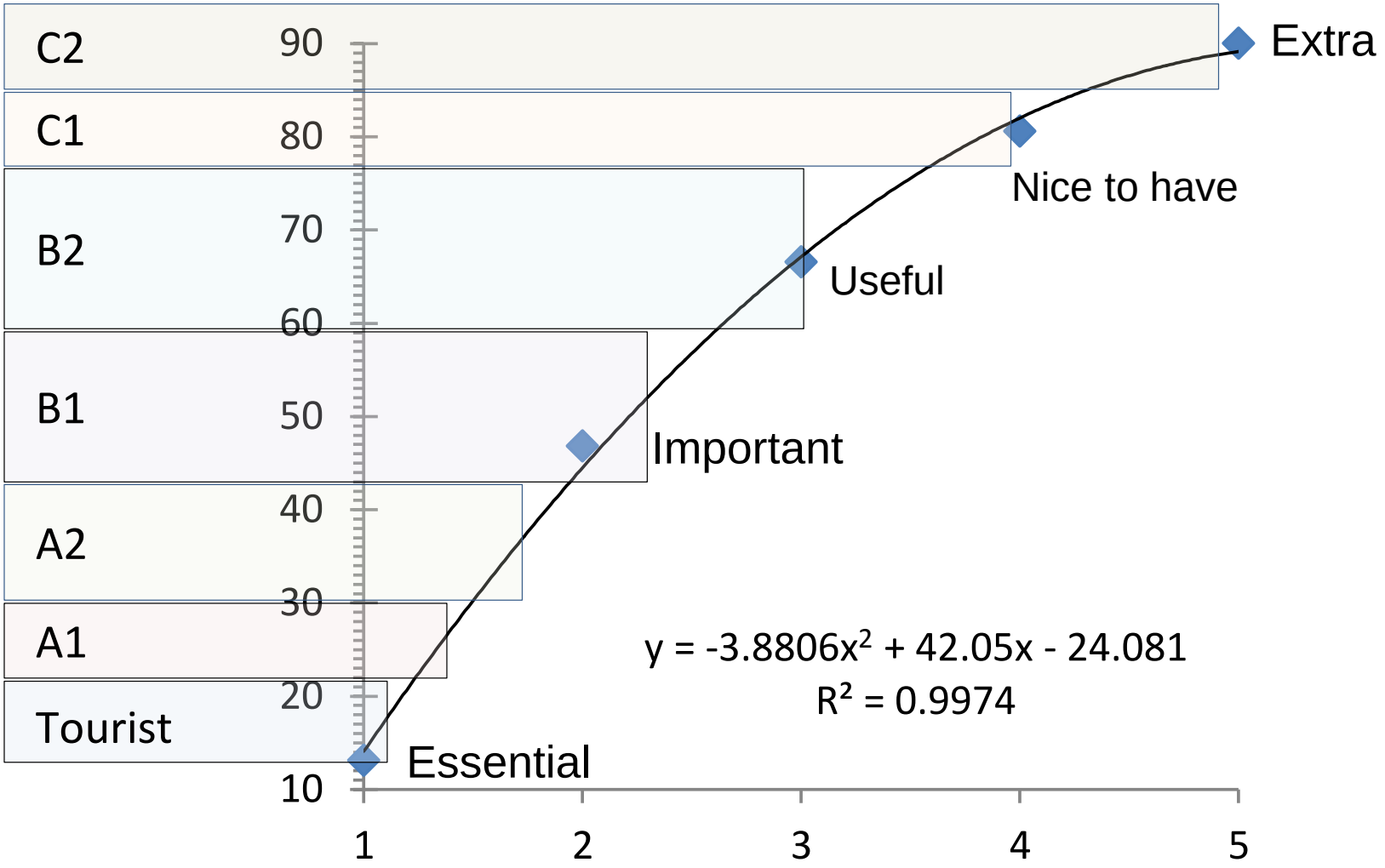


PEARSON

From GSE to VocabScore (1-5)



PEARSON


C2
C1
B2
B1
A2
A1
Tourist
90
80
70
60
50
40
30
20
10
1
2
3
4
5
Extra
Nice to have
Useful
Important
Essential
y = -3.8806x² + 42.05x - 24.081
R² = 0.9974

PEARSON

An example: "FAMILY_*People_personal_traits*"

690 adjectives:
from: *happy* and *good* (A1, 22) to: *sanguine* and *hapless* (C2, 90)

A1: *happy* (22), *good* (22)
A2: *angry* (33), *kind* (36)
A2+: *noisy* (38), *silly* (40)
B1: *helpful* (47), *lonely* (48)
B1+: *nasty* (55), *lively* (56)
B2: *harmless* (59), *mean* (59)
B2+: *grumpy* (68), *thoughtless* (69)
C1: *well-intentioned* (76), *like-minded* (76)
C2: *submissive* (85), *frivolous* (85)

A1 → C2: *quiet* (38) > *calm* (44), *shy* (47)> *timid* (71) > *introverted* (78) > *withdrawn* (83) > *retiring* (85) > *docile* (86) > *demure* (87)

More examples

ball	NN	a round object that you throw, hit, or kick in a game or sport	1463	18	T
ball	NN	something rolled into a round shape	1463	59	B2
ball	NN	a large formal occasion where people dance	1463	69	B2+
ball	NN	the rounded part at the base of the toes or at the base of the thumb	1463	84	C1

bank	NN	company or place where you can keep your money or borrow money	883	18	T
bank	NN	land along the side of a river or lake	883	59	B2
bank	NN	a large sloping mass of earth, sand, snow etc	883	64	B2
bank	NN	a store of something that is kept until people need it	884	70	B2+
bank	NN	a mass of cloud or fog	883	88	C2

Summary & next steps

SUMMARY

- **A probabilistic model** of the relative importance of vocabulary
- Built on the principle of **communicative efficiency in English L1**
- Informing **curriculum design and assessment development**

NEXT STEPS

- **Validation studies** (e.g. comparison with independent measures)
- **Official launch in 2016** (public open access)

Thank you

If you have any question or you would like to hear more about this project and/or **get involved in validation studies**, please contact Veronica Benigno at this email:

veronica.benigno@pearson.com