

**This document is the online-only appendix to:**

## **A new benchmark dataset with production methodology for Short Text Semantic Similarity algorithms**

JAMES O'SHEA, MANCHESTER METROPOLITAN UNIVERSITY  
ZUHAIR BANDAR, MANCHESTER METROPOLITAN UNIVERSITY  
KEELEY CROCKETT, MANCHESTER METROPOLITAN UNIVERSITY

ACM TRANSACTIONS ON SPEECH AND LANGUAGE PROCESSING, VOL. \_\_, NO. \_\_, ARTICLE \_\_, PUBLICATION DATE: \_\_\_\_  
\_\_\_\_\_.

---

### **Contents**

- A. THE FULL DATASET
- B. OTHER STATISTICAL TESTS.
- C. DECOMPOSITION OF ADJECTIVES, VERBS & ADVERBS
- D. POPULATED SAMPLING FRAME
- E. STIMULUS WORD BLOCK STRUCTURE
- F. FULL SET OF CANDIDATE THEMES
- G. COMPOSITE FRAMES COMBINING WORDS AND THEMES
- H. SELECTED SENTENCE PAIRS.
- I. CALL FOR PARTICIPANTS IN A STUDY OF SEMANTIC SIMILARITY
- J. SENTENCE CAPTURE EXPERIMENT INSTRUCTION SHEET
- K. EXAMPLE EXTRACTS FROM QUESTIONNAIRES
- L. PARAPHRASE SENTENCE SUBSET
- M. SAMPLE PARAPHRASE CAPTURE SHEET
- N. CALL FOR PARTICIPANTS IN AN EXPERIMENTAL STUDY OF SENTENCE SEMANTIC SIMILARITY
- O. SURVEY: SENTENCE SEMANTIC SIMILARITY
  - SAMPLE CARDS WITH SENTENCE PAIRS
  - SIMILARITY RATING SHEET (sheet 1 of 2)
  - PARTICIPANT DETAILS (Student)

## A. THE FULL DATASET

The first column is the number of the sentence pair. The second is the two sentences making up the pair. The third is the semantic similarity rating calculated as the average of the human ratings for the sentence pair (0.00 – 4.00). The final column is the standard deviation of the human ratings, which gives a measure of noisiness. The two faint entries are calibration pairs borrowed from STSS-65. These should NOT be used in calculations and are for reference only. For a fuller understanding of how the data set was collected and the method for using it to compare STSS measures please see [O'Shea, et al. 2008] which you may also wish to cite. For complete details of the methodology involved see [O'Shea 2010].

Table 1. semantic similarity ratings for STSS-131 (on a scale from 0.00 to 4.00)

SP	Sentences	$\bar{X}$ (Human ratings of semantic similarity)	$S$
66	Would you like to go out to drink with me tonight? I really don't know what to eat tonight so I might go out somewhere.	1.01	0.77
67	I advise you to treat this matter very seriously as it is vital. You must take this most seriously, it will affect you.	3.38	0.69
68	When I was going out to meet my friends there was a delay at the train station. The train operator announced to the passengers that the train would be delayed.	3.13	0.68
69	Does music help you to relax, or does it distract you too much? Does this sponge look wet or dry to you?	0.1	0.29
70	You must realise that you will definitely be punished if you play with the alarm. He will be harshly punished for setting the fire alarm off.	2.84	0.87
71	I will make you laugh so much that your sides ache. When I tell you this you will split your sides laughing.	3.75	0.38
72	You shouldn't be covering what you really feel. There is no point in covering up what you said, we all know.	2.21	0.97
73	Do you want to come with us to the pub behind the hill? We are going out for drinks tonight in Salford Quays if you would like to come.	1.82	1.09
74	This key doesn't seem to be working, could you give me another? I dislike the word quay, it confuses me, I always think of things for locks, there's another one.	0.72	0.87
75	The ghost appeared from nowhere and frightened the old man. The ghost of Queen Victoria appears to me every night, I don't know why, I don't even like the royals.	1.45	0.75
76	You're not a good friend if you're not prepared to be present when I need you. A good friend always seems to be present when you need them.	3.14	0.94
77	The children crossed the road very safely thanks to the help of the lollipop lady. It was feared that the child might not recover, because he was seriously ill.	0.13	0.29
78	I have invited a variety of people to my party so it should be interesting. A number of invitations were given out to a variety of people inviting them down the pub.	2.18	0.88
79	I offer my condolences to the parents of John Smith, who was unfortunately murdered. I express my sympathy to John Smith's parents following his murder.	3.91	0.23
80	Boats come in all shapes and sizes but they all do the same thing. Chairs can be comfy and not comfy, depending on the chair.	0.5	0.69
81	If you continuously use these products, I guarantee you will look very young. I assure you that, by using these products consistently over a long period of time, you will appear really young.	3.58	0.57
82	We ran farther than the other children that day.	2.43	1.06

	You ran farther than anyone today.		
83	I always like to have a slice of lemon in my drink especially if it's Coke. I like to put a wedge of lemon in my drinks, especially cola.	3.81	0.55
84	It seems like I've got eczema on my ear doctor, can you recommend something for me? I had to go to a chemist for a special rash cream for my ear.	2.05	0.9
85	I am proud of our nation, well, most of it. I think of myself as being part of a nation.	1.71	1.03
86	There was a heap of rubble left by the builders outside my house this morning. Sometimes in a large crowd accidents may happen, which can cause deadly injuries.	0.09	0.27
87	Water freezes at a certain temperature, which is zero degrees Celsius. The temperature of boiling water is 100 C and the temperature of ice is 0 C.	3.08	0.98
88	We got home safely in the end, although it was a long journey. Though it took many hours travel, we finally reached our house safely.	3.06	0.95
89	A man called Dave gave his fiancée a large diamond ring for their engagement. The man presented a diamond to the woman and asked her to marry him.	3.22	0.73
90	I used to run quite a lot, in fact once I ran for North Tyneside. I used to climb lots at school as we had a new climbing wall put in the gym.	0.74	0.75
91	I love to laugh as it makes me happy as well as those around me. I thought we bargained that it would only cost me a pound.	0.08	0.32
92	Because I am the eldest one I should be more responsible. Just because of my age, people shouldn't think I'm a responsible adult, but they do?	2.23	0.79
93	I need to dash into the kitchen because I think my chip pan is on fire. In the event of a chip pan fire follow the instructions on the safety note.	1.7	1.03
94	Peter was a very large youth, whose size intimidated most people, much to his delight. Now I wouldn't say he was fat, but I'd certainly say he was one of the larger boys.	1.96	0.95
95	I'm going to buy a grey jumper today, in half an hour. That's a nice grey top, where did you get it from?	1.25	0.98
96	We got soaked in the rain today, but now we are nice and dry. I was absolutely soaking wet last night, I drove my bike through the worst weather.	1.68	0.75
97	Global warming is what everyone is worrying about today. The problem of global warming is a concern to every country in the world at the moment.	3.14	0.84
98	He was harshly punished for setting the fire alarms off. He delayed his response, in order to create a tense atmosphere.	0.22	0.6
99	Midday is 12 o'clock in the middle of the day. Noon is 12 o'clock in the middle of the day.	3.96	0.16
100	That's not a very good car, on the other hand mine is great. This is a terrible noise level for a new car.	1.05	0.95
101	There was a terrible accident, a pileup, on the M16 today. It was a terrible accident, no one believed it was possible.	2.33	0.93
102	After hours of getting lost we eventually arrived at the hotel. After walking against the strong wind for hours he finally returned home safely.	1.09	0.91
103	The first thing I do in a morning is make myself a cup of coffee. The first thing I do in the morning is have a cup of coffee.	3.85	0.39
104	Someone spilt a drink accidentally on my shirt, so I changed it. It appears to have shrunk, it wasn't that size before I washed it.	0.48	0.72
105	I'm worried most seriously about the presentation, not the essay. It is mostly very difficult to gain full marks in today's exam.	0.77	0.82
106	It is mostly very difficult to gain full marks in today's exam. The exam was really difficult, I've got no idea if I'm going to pass.	2.54	0.98

107	Meet me on the hill behind the church in half an hour. Join me on the hill at the back of the church in thirty minutes time.	3.93	0.25
108	If you don't console with a friend, there is a chance you may hurt their feelings. One of the qualities of a good friend is the ability to console.	3.01	0.85
109	We tried to bargain with him but it made no difference, he still didn't change his mind. I tried bargaining with him, but he just wouldn't listen.	3.43	0.54
110	It gives me great pleasure to announce the winner of this year's beauty pageant. It's a real pleasure to tell you who has won our annual beauty parade.	3.88	0.24
111	They said they were hoping to go to America on holiday. I like to cover myself up in lots of layers, I don't like the cold.	0.16	0.5
112	Will I have to drive far to get to the nearest petrol station? Is it much farther for me to drive to the next gas station?	3.84	0.37
113	I think I know her from somewhere because she has a familiar face. You have a very familiar face, where do I know you from?	3.36	0.8
114	I am sorry but I can't go out as I have a heap of work to do. I've a heap of things to finish so I can't go out I'm afraid.	3.6	0.72
115	The responsible man felt very guilty when he crashed into the back of someone's car. A slow driver can be annoying even though they are driving safely.	0.88	0.75
116	Get that wet dog off my brand new white sofa. Make that wet hound get off my white couch – I only just bought it.	3.59	0.86
117	He fought in the war in Iraq before being killed in a car crash. The prejudice I suffered whilst on holiday in Iraq was quite alarming.	0.55	0.65
118	The cat was hungry so he went into the back garden to find lunch. The hen walked about in the yard eating tasty grain.	1.2	0.82
119	My bedroom wall is lemon coloured but my mother says it is yellow. Roses can be different colours, it has to be said red is the best though.	0.68	0.77
120	Would you like to drink this wine with your meal? Will you drink a glass of wine while you eat?	3.56	0.65
121	Roses can be different colours, it has to be said red is the best though. Roses come in many varieties and colours, but yellow is my favourite.	2.83	0.9
122	Flies can also carry a lot of disease and cause maggots. I dry my hair after I wash it or I will get ill.	0.12	0.28
123	Could you climb up the tree and save my cat from jumping please? Can you get up that tree and rescue my cat otherwise it might jump?	3.83	0.34
124	The pleasure that I get from studying, is that I learn new things. I have a doubt about this exam, we never got to study for it.	0.74	0.76
125	The perpetrators of war crimes are rotten to the core. There are many global issues that everybody should be aware of, such as the threat of terrorism.	0.95	0.91
126	The damp was mostly in the very corner of the room. The young lady was somewhat partially burnt from the sun.	0.11	0.31
127	We often ran to school because we were always late. I knew I was late for my class so I ran all the way to school.	3.1	0.85
128	I hope you're taking this seriously, if not you can get out of here. The difficult course meant that only the strong would survive.	0.5	0.87
129	The shores or shore of a sea, lake or wide river is the land along the edge of it. An autograph is the signature of someone famous which is specially written for a fan to keep.	0.11	0.43
130	I bought a new guitar today, do you like it? The weapon choice reflects the personality of the carrier.	0.16	0.34
131	I am so hungry I could eat a whole horse plus dessert. I could have eaten another meal, I'm still starving.	3.06	0.85

The following guidance is intended to help make benchmark tests performed with the data set comparable.

1. An STSS measure can be validated by comparing its performance with human ratings, in particular the ratings that a “typical” human might give.

2. The ratings in the table follow the practice used in word similarity studies [Miller and Charles 1991]. The “typical” human rating is the mean of those given by a set of participants. The measure of agreement is the Pearson product-moment correlation coefficient ( $r$ ) quoted with statistical significance. The final column,  $S$ , is the corresponding standard deviation for each mean, a measure of noisiness or lack of precision of the ratings.

3. The ratings are the numbers in the  $\bar{X}$  column; they are from a rating scale running from 0.00 to 4.00. The simplest procedure is to calculate the correlation coefficient between a new measure and the human ratings in the original range (0.00 – 4.00). Linear transformations are permissible, e.g. dividing by 4 to re-scale them to run from 0.00 to +1.00. Re-scaling should not lead to a different correlation co-efficient (however, see below on rounding noise).

4. Consistency with other studies. Most STSS algorithms produce measures in the range from 0 to +1. Applying different rounding procedures can introduce noise and lead to variations in the least significant digit of  $r$ . For consistency with other studies, round the ratings from the STSS algorithm to 3 decimal places. Then calculate  $r$ , and round  $r$  to 3 decimal places. Common sense dictates that as the least significant digit of the 3 is based on the estimated digit, the importance of differences between measures based on this digit alone should not be exaggerated.

5. Those familiar with measurement theory may argue that mean and  $r$  are unsuitable statistics for data collected on this measurement scale. We are aware of the argument; however we have used the techniques because they are well-established and understood in the field of word similarity. Furthermore in the data collection process and the steps taken to improve ratio scale properties are described in detail in [O’Shea, et al. 2008] and [O’Shea 2010] as well as the current paper.

6. The calibration sentence pairs (SP99 and SP129) are taken from STSS-65 and **should not be used as part of this dataset.**

## B. OTHER STATISTICAL TESTS.

Various tests for significance are appropriate in different circumstances. If we want to test the statistical significance of the difference between one STSS algorithm and another, these are dependent samples and the appropriate test is Steiger's z-test. This requires the construction of a correlation triangle, described later. The one-sample t-test can be used to compare a single correlation coefficient with an average of correlation coefficients (e.g. STASIS with the STSS-131 average from human raters). The 2-sample t-test can be used to compare averages from independent samples (e.g. to find a significant difference between the STSS-65 and STSS-131 datasets). Finally Fisher's r-to-z test can be used to compare correlation coefficients for the same algorithm across two different datasets (e.g. difference between LSA on STSS-65 vs. STSS-131).

### USING STEIGER'S Z-TEST TO COMPARE TWO CORRELATION COEFFICIENTS (DEPENDENT SAMPLES)

Using Steiger's test to compare 2 correlation coefficients requires the construction of a correlation triangle. For example, consider comparing the correlation between STASIS and STSS-131 human ratings with the correlation between LSA and STSS-131 human ratings. Correlation triangles are formed according to Figure 1 and the specific triangle required for this calculation is shown in Figure 2.

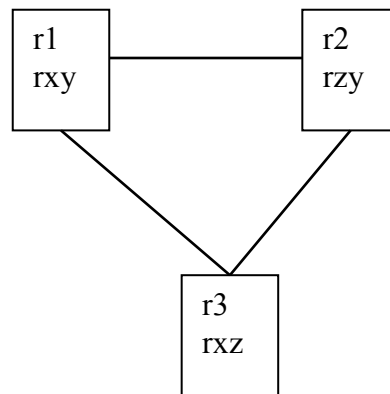


Fig. 1 General form of correlation triangle

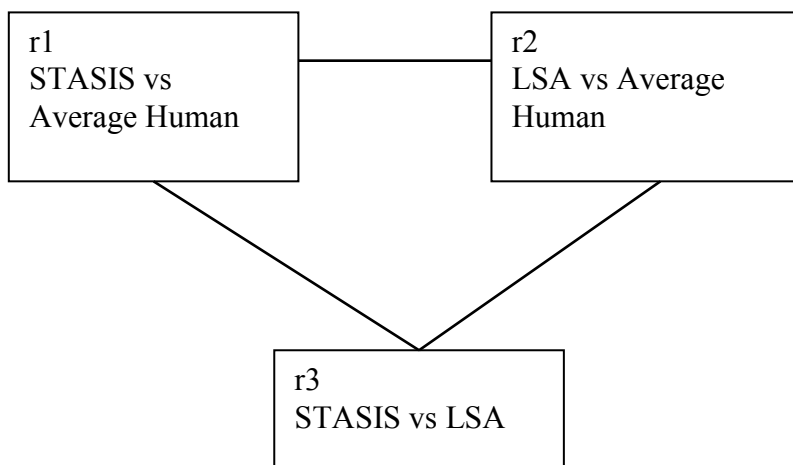


Fig. 2 Specific correlation triangle for STASIS vs LSA

From Table IV in the main paper:

r1	rx	STASIS vs Average humans	0.636
r2	ry	LSA vs Average human	0.693
n=64 (64 Sentence Pairs without the two calibration pairs)			

Calculated correlation:

r3	rxz	STASIS vs LSA	0.52
----	-----	---------------	------

Applying the test gives the following results:

z-values for all differences:  
Method Steigers Z

$0.636 - 0.693 = -0.057$ ;  $z = -0.677$ ;  $p = 0.7507$   
(left p: 0.2493; two sided: 0.4986)

$0.636 - 0.52 = 0.116$ ;  $z = 1.48$ ;  $p = 0.0695$   
(left p: 0.9305; two sided: 0.139)

$0.693 - 0.52 = 0.173$ ;  $z = 2.126$ ;  $p = 0.0167$   
(left p: 0.9833; two sided: 0.0334)

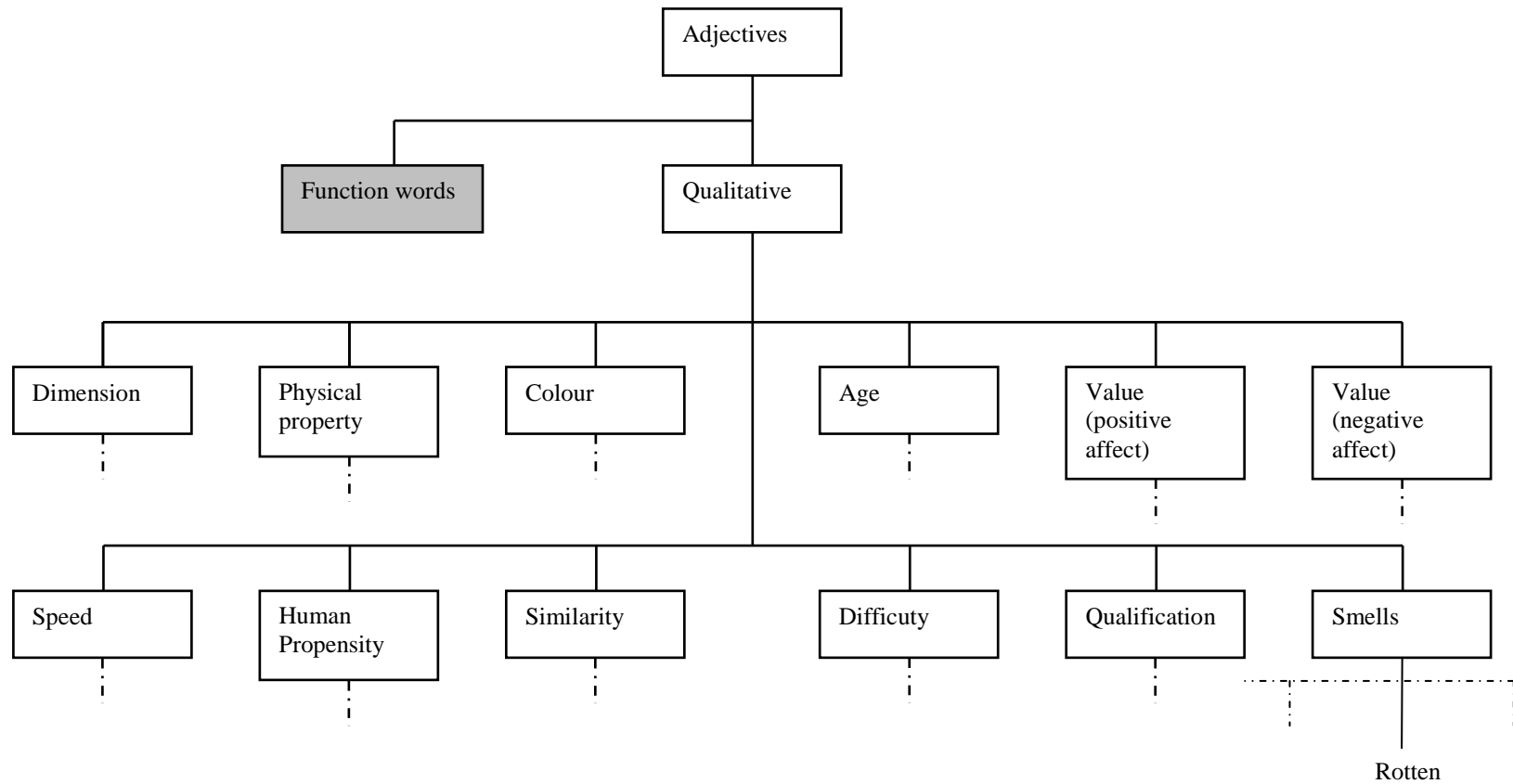
At the time of writing, Steiger's test was supported in neither Minitab nor SPSS. However, online calculators were available and two different calculators were used which gave consistent results, with very small differences attributable to operations such as rounding within floating point calculations. The calculators were available at:

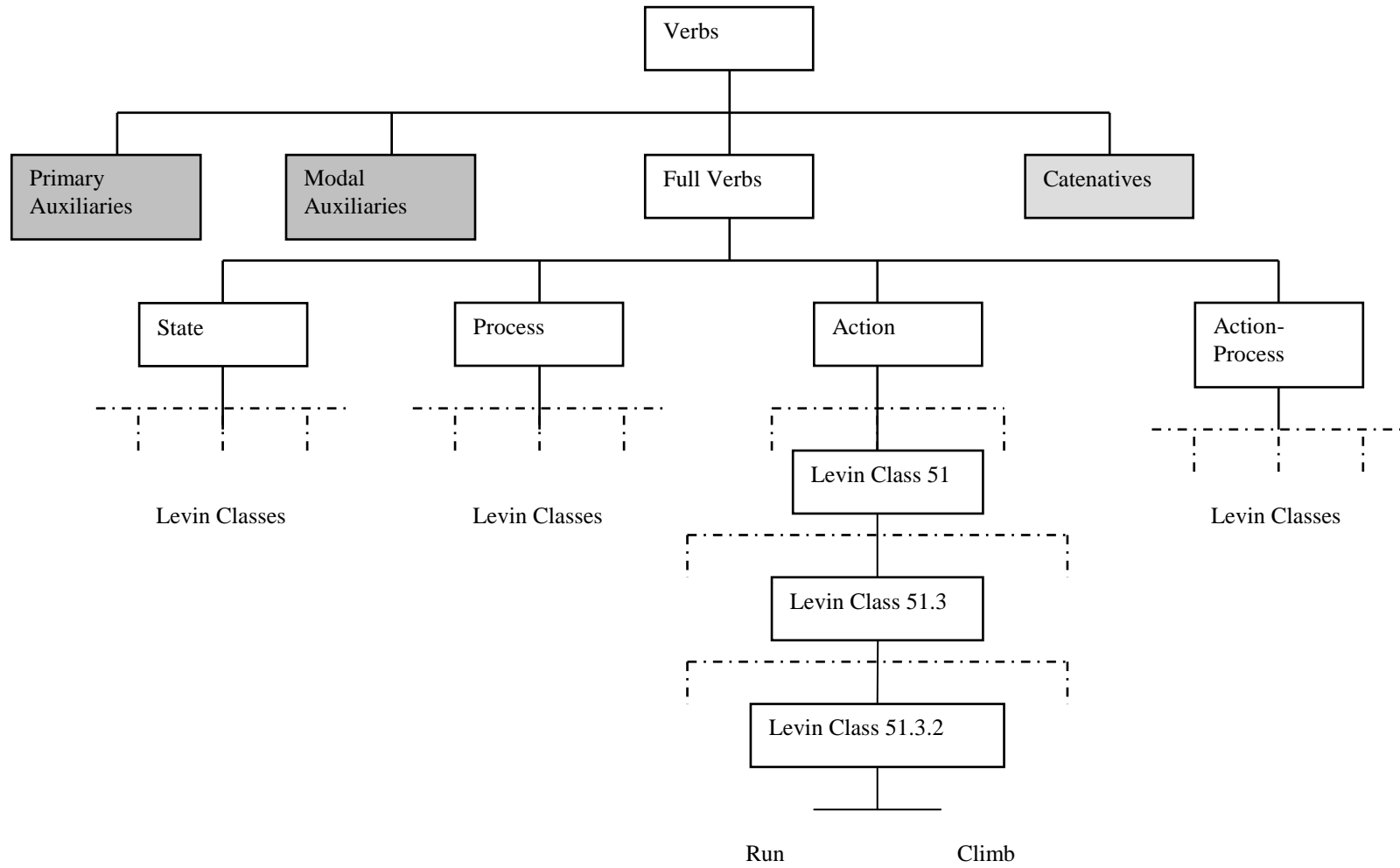
[Uitenbroek 2013] and [Grabin 2013]

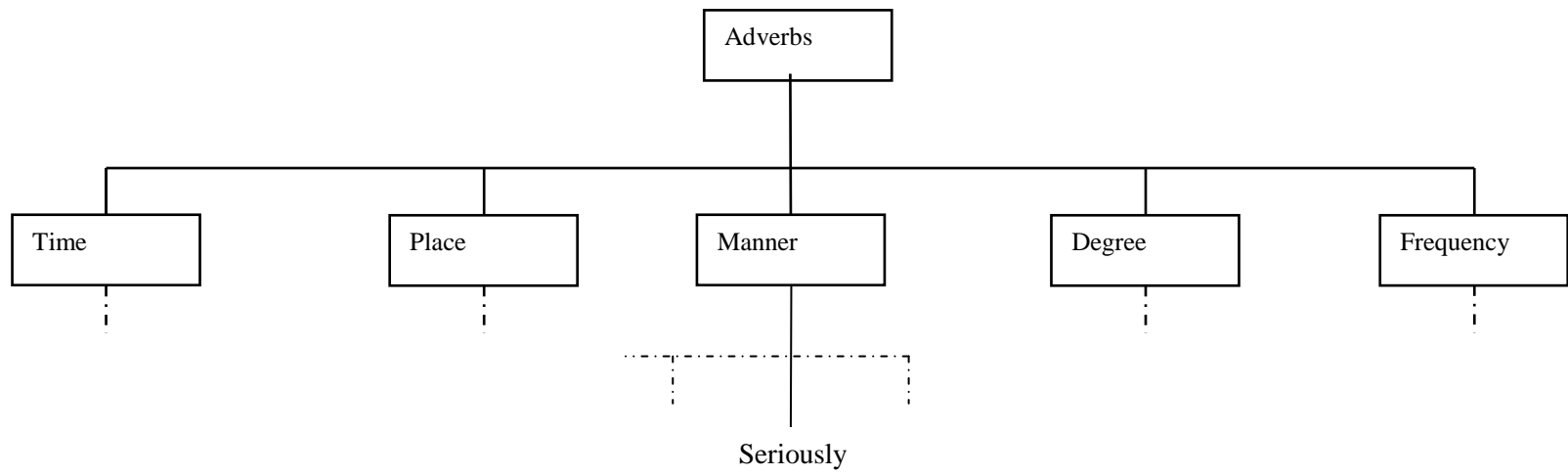
Note that the test compares single correlation coefficients, not a correlation against an average of correlation coefficients, so in this case for "Average Human" we used the human with the inter-rater agreement closest to the average inter-rater agreement.



**C. DECOMPOSITION OF ADJECTIVES, VERBS & ADVERBS**







#### D. POPULATED SAMPLING FRAME

No.	Class	Word	Additional Criteria / comments
1	Noun abstract quality	Variety	
2	Noun abstract idea from science	Temperature	
3	Noun abstract idea institutional fact	Nation	
4	Noun abstract idea	Delay	Homonymous noun-verb pair LF
5	Noun abstract state	Conflict	
6	Noun abstract emotion positive	Pleasure	
7	Noun abstract emotion negative	Doubt	
8	Noun abstract	Prejudice	Randomly selected LF
9	Noun concrete living biological body parts	Ear	
10	Noun concrete living biological fruits and vegetables	Lemon	LF
11	Noun concrete living biological animals	Cat	
12	Noun concrete living biological birds	Hen	LF
13	Noun concrete living biological insects	Fly	
14	Noun concrete living biological plants and flowers	Rose	LF
15	Noun concrete nonliving biological foodstuffs	Coffee	
16	Noun concrete nonliving nonbiological clothing	Shoe	*
17	Noun concrete nonliving nonbiological tools and Manipulables	Key	Source for homophone
18	Noun concrete nonliving nonbiological vehicles	Boat	Normally found outdoors
19	Noun concrete nonliving nonbiological furniture	Chair	Normally found indoors
20	Noun concrete nonliving nonbiological musical instruments	Guitar	LF
21	Noun concrete nonliving nonbiological miscellaneous artefacts	Weapon	
22	Noun concrete nonliving nonbiological gemstones	Diamond	LF
23	Noun concrete nonliving nonbiological other non man-made	Hill	
24	Noun collective living biological	Crowd	

No.	Class	Word	Additional Criteria / comments
25	Noun collective nonliving nonbiological	Heap	LF
26	Noun concrete living biological	Parent	Randomly selected
27	Noun concrete nonliving nonbiological	Fire	Randomly selected
28	Noun Homonym/homophone	Quay	Homophone of Key
29	Adjective dimension	Large	Source for comparative
30	Adjective physical property	Dry	Source for antonym, wet : Homonymous verb-adjective pair
31	Adjective colour	Grey	
32	Adjective age	Young	
33	Adjective Value positive	Great	NOT an antonym of terrible
34	Adjective Value negative	Terrible	NOT an antonym of great
35	Adjective speed	Slow	
36	Adjective human propensity	Responsible	
37	Adjective similarity	Familiar	
38	Adjective difficulty	Difficult	
39	Adjective qualification	Global	LF
40	Adjective smell	Rotten	LF
41	Adjective comparative	Larger	Comparative of large
42	Adjective antonym	Wet	Antonym of dry
43	Verb State (state experiential)	Hope	Levin class 32.2 long (for)
44	Verb State (state locative, continuous locative)	Cover	Levin classes 47.8 contiguous location (also 9.8 fill)
45	Verb State	Relax	Levin class 31.1 amuse LF
46	Verb Process	Change	Levin classes 41.1.1 dress (also 45.4 CoS, 26.6 turn, 13.6 exchange)
47	Verb Process	Appear	Levin class 48.1 (48.1.1 appear)
48	Verb Process	Dry	Levin class 45.4 other change of state LF : Homonymous verb-adjective pair
49	Verb Action	Run	Levin classes 51.3.2 run (also 26.3 preparing, 47.5.1 swarm, 47.7 meander) Source for Levin 3 <sup>rd</sup> level class pair
50	Verb Action	Laugh	Levin class 40.2 nonverbal expression
51	Verb Action	Bargain	Levin class 36.1 correspond
52	Verb Action-Process	Drink	Levin class 39.1 ingesting Source for Levin 2 <sup>nd</sup> level class pair
53	Verb Action-Process	Punish	Levin class 33 judgement (negative)
54	Verb Action-Process	Delay	Levin class 53.1 lingering Homonymous noun-verb pair LF
55	Verb Action-Process	Eat	Levin class 39.1 ingesting Paired with drink

No.	Class	Word	Additional Criteria / comments
56	Verb Action	Climb	Levin class 51.3.2 run Paired with run
57	Adverb Time	Eventually	
58	Adverb Place	Far	
59	Adverb Manner	Seriously	
60	Adverb Degree	Partially	LF
61	Adverb Frequency	Mostly	LF
62	Adverb	Safely	Randomly selected
63	Adverb Comparative	Farther	
64	Adverb Superlative	Most seriously	

\* Due to a minor procedural error, sentences stimulated by “shoe” did not qualify for inclusion. The class Noun:concrete:nonliving:nonbiological:clothing is successfully represented in the set by sentence pair 95, where sentences contain the word “jumper” and “top.”

## E. STIMULUS WORD BLOCK STRUCTURE

Block A1	Block A2	Block B1	Block B2
7 Nouns variety (AQu) conflict (ASt) ear (LBBoP) cat (LBA) weapon (NNMiAr) heap (CNNLF) key (NNTMHe/Ho)	7 Nouns delay (ARLFPoV) pleasure (AEm+) lemon (BFrV) parent (LBR) coffee(NBFo) boat (NNVeOut) diamond (NNGLF)	7 Nouns nation (AIF) doubt (AEm-) rose (BPFILF) fly (LBI) chair (NNFuIn) hill (NNONMM) fire(NNRPoVb)	7 Nouns temperature (ASc) prejudice (ARLF) hen (LBBiLF) shoe (NNCl) guitar (NNMuILF) crowd (CLB) quay (NNMiArHoHo)
4 Verbs hope 32.2 (s) laugh 40.2 (a) relax 31.1 (s) (lf) climb 51.3.2 (a)(SSC)	3 Verbs change 41.1.1 (p) drink 39.1 (ap)(SC) punish 33 (ap)	4 Verbs cover 47.8 (s) run 51.3.2 (a)(SSC) bargain 36.1 (a) (lf) eat 39.1 (ap)(SC)	3 Verbs dry 45.4 (p) (lf) (PoA) delay 53.1 (ap) (lf) (PoN) appear 48.1.1 (p)
3 Adjectives large great (V+) difficult	4 Adjectives dry (AntPoV) slow global (LF) young	3 Adjectives wet (Ant) terrible (V-) familiar	4 Adjectives rotten (LF) responsible grey larger
2 Adverbs seriously farther (further) (Cmp)	2 Adverbs far mostly	2 Adverbs eventually most seriously (Sup)	2 Adverbs safely partially

The blocked design distributes the workload of sentence production amongst participants whilst avoiding spurious semantic overlap. A participant receives a work package based on one of the four blocks (A1, A2, B1 or B2).

Colour coding allows the distribution of word properties amongst the participants to be monitored. In particular **WORDS IN RED** must not be moved to other columns without checking side effects, the codes PoV, PoN, PoA indicate variants of homonymous pairings of the same word (e.g. delay) in different classes, where both members of the pair should not be presented to the same participant.

### Key

A / C – Abstract / Collective (concrete not labeled)

LF – Low Frequency

L / N – Living / Non-Living

R – Randomly selected

B / N – Biological / Non-biological

PoV / PoN / PoA – Polysemous Verb / Polysemous Noun / Polysemous adjective pairing

## F. FULL SET OF CANDIDATE THEMES

Theme	KS3G	TYE	TYR	NYT
The School (subjects, timetable preferences etc.)	Y	N	N	N
Greeting people and talking about yourself, taking leave	Y	N	Y	Y
Likes & dislikes, expressing opinions	Y	N	Y	Y
Activities, pastimes, sports, clubs, talents	Y	N	N	Y
Shopping & money	Y	Y	Y	Y
Where are you from, background	Y	N	N	Y
About the house (location & movement), household	Y	Y	N	Y
Describing family, relatives, people	Y	Y	Y	Y
Eating out (restaurant etc.)	Y	Y	Y	Y
Directions, tourism, in the street, around the town, where you are going	Y	N	Y	Y
Daily routines (& telling time)	Y	N	Y	Y
Describing people	Y	Y	Y	Y
Travelling and getting about, public transport	Y	N	Y	Y
Dealing with money (transactions, earning, shopping)	Y	N	Y	Y
Pets, animals	Y	Y	N	N
Clothes	Y	Y	N	N
Holidays, Hotels, Airports	Y	Y	Y	Y
Shopping for food & drink, drink, ordering drinks	Y	Y	Y	Y
Work & Jobs	Y	Y	N	Y
The Weather	Y	Y	Y	N
Future Events, speaking about the future	Y	N	N	Y
Past Events, saying what you did	Y	N	N	Y
Body & Health, doctor's surgery, feeling ill, injuries	Y	Y	Y	Y
Using the car	Y	N	N	N
Writing a letter	Y	N	N	N
Music (instruments, concert, pop)	Y	N	N	N
Getting help in an emergency	Y	N	N	N
Going out (socially), giving invitations	Y	N	N	Y
Talking about magazines	Y	N	N	N
Complaints & problems	N	N	Y	N
At the post office	N	N	Y	N
Discussing languages	N	N	N	Y
Giving orders and instructions	N	N	N	Y
Giving invitations	N	N	N	Y
Asking permission, for favours	N	N	N	Y
Exchanging news	N	N	N	Y
Living and working abroad	N	N	N	Y

KS3G = Key Stage 3 German, the target attainment 4 of key stage 3 is concerned with the production of texts containing 2 or 3 short sentences on familiar topics [Hawkin, T 1995]

TYE = Teach Yourself English [Hunt 2001]

TYR = Teach Yourself Russian [Farmer 1996]

NYT = Now Your Talking (a multimedia Irish course for beginners) [O'Donaill 1995]



## G. COMPOSITE FRAMES COMBINING WORDS AND THEMES

Block A1 Nouns	Block A2 Nouns	Block B1 Nouns	Block B2 Nouns
Living/Bio body parts <b>Ear HF</b>	Living/Bio body parts <b>Hand HF</b> 12 Holiday time & travel 12.5 2E Services <b>consulting a doctor, dentist or chemist</b>	Living/Bio body parts <b>Eye HF</b> 12 Holiday time & travel 12.5 2E Services <b>consulting a doctor, dentist or chemist</b>	Living/Bio body parts <b>Nose HF</b> 12 Holiday time & travel 12.5 2E Services <b>consulting a doctor, dentist or chemist</b>
Nl/Nbio found indoors <b>Desk HF</b>		NL/Nbio found indoors <b>Chair HF</b>	
Block A1 Verbs 48.1.2 <b>present LF</b> 14 Theme 4. The Young Person in Society 14.1 4A Character and Personal Relationships <b>the qualities of a good friend or relationship</b>	Block A2 Verbs 31.1 <b>console VLF</b> 14 Theme 4. The Young Person in Society 14.1 4A Character and Personal Relationships <b>the qualities of a good friend or relationship</b>	Block B1 Verbs 31.1 <b>calm LF</b> 14 Theme 4. The Young Person in Society 14.1 4A Character and Personal Relationships <b>the qualities of a good friend or relationship</b>	Block B2 Verbs 31.1 <b>comfort LF</b> 14 Theme 4. The Young Person in Society 14.1 4A Character and Personal Relationships <b>the qualities of a good friend or relationship</b>
Block A1 Adjectives Colour <b>Green HF</b>	Block A2 Adjectives Colour <b>White HF</b>	Block B1 Adjectives Colour <b>Yellow LF</b>	<b>Block B2 Adjectives</b> Colour <b>Grey HF</b>
			Smell <b>Stale VLF</b>
Block A1 Adverbs Time <b>Already HF</b>	Block A2 Adverbs Time <b>Finally HF</b>	<b>Block B1 Adverbs</b> Time <b>Eventually HF</b>	Block B2 Adverbs Time <b>Still HF</b>
		Place <b>Across LF</b>	

These were used to create supplementary materials.. In the first row, participants in blocks A2, B1 and B2 who do not see the word “ear” are presented with different words from the same category (e.g. hand) and asked to use it in a sentence on the general topic of “consulting a doctor, dentist or chemist.” The corresponding work for the block A1 participant is to write a sentence using the word “desk” which corresponds with the word “chair” in block B1.

The other pairs of rows are similar, with the exception of the verbs row, This combined a new theme with the opportunity to use new verbs from level 2 Levin classes which had already been used (31.1 relax. 48.1 appear).

Both frames were used and contributed small quantities of usable additional data, as well as preventing the participants from rushing the task. They offered the opportunity to collect more candidates for medium-high similarity sentence pairs.

Block A1 Nouns	Block A2 Nouns	Block B1 Nouns	Block B2 Nouns
<b>Conflict</b>	<b>Disagreement</b> (A1 conflict)	<b>War</b> (A1 conflict)	<b>Agreement</b> (A1 conflict)
<b>Valley</b> (B1 Hill)			
Block A1 Verbs <b>Dash</b> (Run B1)	Block A2 Verbs <b>Gush</b> (Run B1)	Block B1 Verbs <b>Run</b>	Block B2 Verbs <b>Manage</b> (Run B1)
		Swallow (Drink A2)	
Block A1 Adjectives <b>Childish</b> (Young A2) 11. My world 11.1.A Exchange information about self, family and friends	Block A2 Adjectives <b>Young</b>	Block B1 Adjectives <b>New</b> (Young A2) 11. My world 11.1.A Exchange information about self, family and friends	Block B2 Adjectives <b>Fresh</b> (Young A2) 11. My world 11.1.A Exchange information about self, family and friends
	<b>Famous</b> (Great A1) 11 My World 11.3 1C Home and Local Give and seek description of your/other.s town, neighbourhood		
Block A1 Adverbs <b>Securely</b> (Safely B2) 12 Holiday Time & Travel 12.5 2E Services Exchange information about a loss or theft	Block A2 Adverbs <b>Carefully</b> (Safely B2) 12 Holiday Time & Travel 12.5 2E Services Exchange information about a loss or theft	Block B1 Adverbs <b>Carelessly</b> (Safely B2) 12 Holiday Time & Travel 12.5 2E Services Exchange information about a loss or theft	Block B2 Adverbs <b>Safely</b>
<b>Seriously</b>			Block B2 Adverbs <b>Slightly</b> (A1 Seriously) 12.5 2E Services 12 Holiday Time & Travel Exchange information about a loss or theft

**H. SELECTED SENTENCE PAIRS. The sentence pairs extracted from the database**

The table shows the selected sentence pairs, with their predicted similarity (by the judges). Red sentence pairs are those where a low similarity pair was removed to balance the distribution after the pilot experiment by adding high similarity pairs (paraphrases). Criteria are those used in constructing the sampling frame (plus miscellaneous properties) from section 3 of the paper. The sentences are in blocks running from predicted high to low similarity.

High – 15

Stimulus word pair	Pred Sim	dB entry	Sentence Pair	Criteria / Properties
Seriously Most seriously		138	I advise you to treat this matter very seriously as it is vital.	Adverb-Adverb Superlative Instruction
	H	955	You must take this most seriously, it will affect you.	
laugh		1024	I will make you laugh so much that your sides ache.	To meet high similarity target Commitment Paraphrased sentence (SP71)
			When I tell you this you will split your sides laughing.	
Hill-Quay Going_out		184	Do you want to come with us to the pub behind the hill?	Concrete Noun-Noun To meet high similarity target Same theme / different word
	H	423	We are going out for drinks tonight in Salford Quays if you would like to come.	
parent		1088	I offer my condolences to the parents of John Smith, who was unfortunately murdered.	Expression Paraphrased sentence (SP79)
			I express my sympathy to John Smith's parents following his murder.	
Lemon		581	I always like to have a slice of lemon in my drink especially if it's Coke.	To meet high similarity target Noun Low Frequency Living Biological Statement
	H	141	I like to put a wedge of lemon in my drinks, especially cola.	
Diamond		276	A man called Dave gave his fiancée a large diamond ring for their engagement.	Noun Non-Living/Non-Biological

	H	451	The man presented a diamond to the woman and asked her to marry him.	To meet high similarity target Low Frequency Statement
Global		471	Global warming is what everyone is worrying about today.	To meet high similarity target Low Frequency Adjective Statement
	H	381	The problem of global warming is a concern to every country in the world at the moment.	
Terrible		507	There was a terrible accident, a pileup, on the M16 today.	To meet high similarity target Adjective Value negative
	H	940	It was a terrible accident, no one believed it was possible.	
Coffee		575	The first thing I do in a morning is make myself a cup of coffee	To meet high similarity target Non-Living/Biological Statement
	H	817	The first thing I do in the morning is have a cup of coffee.	
Run		194	We often ran to school because we were always late.	To meet high similarity target Verb
	H	986	I knew I was late for my class so I ran all the way to school.	
Bargain		80	We tried to bargain with him but it made no difference, he still didn't change his mind.	To meet high similarity target Verb Statement
	H	952	I tried bargaining with him, but he just wouldn't listen.	
Familiar		76	I think I know her from somewhere because she has a familiar face.	Adjective To meet high similarity target
	H	942	You have a very familiar face, where do I know you from?	
Eat		192	I am so hungry I could eat a whole horse plus dessert.	Verb To meet high similarity target
	H	946	I could have eaten another meal, I'm still starving.	
Punish		1123	You must realise that you will definitely be punished if you play with the alarm.	Verb To meet high similarity target
	H	1124	He will be harshly punished for setting the fire alarm off.	
Rose			Roses can be different colours, it has to be said red is the best though.	Noun To meet high similarity target
	H	1125	Roses come in many varieties and colours , but yellow is my favourite.	

Selected Sentence Pairs

High-Medium - 5

<b>Present Good_friend</b>		<b>1038</b>	<b>You're not a good friend if you're not prepared to be present when I need you.</b>	<b>To meet similarity target Verb</b>
	<b>H- M</b>	<b>764</b>	<b>A good friend always seems to be present when you need them.</b>	<b>Same word / same theme Statement</b>
<b>Variety</b>		<b>62</b>	<b>I have invited a variety of people to my party so it should be interesting.</b>	<b>To meet high similarity target Adjective</b>
	<b>H- M</b>	<b>741</b>	<b>A number of invitations were given out to a variety of people inviting them down the pub.</b>	<b>Same word / same theme</b>
<b>Farther</b>		<b>757</b>	<b>We ran farther than the other children that day.</b>	<b>To meet similarity target Adverb</b>
	<b>H- M</b>	<b>563</b>	<b>You ran farther than anyone today.</b>	<b>Same word</b>
<b>Ear Doctor_dentist</b>		<b>603</b>	<b>It seems like I've got eczma on my ear doctor, can you recommend something for me?</b>	<b>To meet similarity target Noun Ear</b>
		<b>485</b>	<b>I had to go to a chemist for a special rash cream for my ear.</b>	<b>Same word (SP84)</b>
<b>Safely</b>		<b>677</b>	<b>We got home safely in the end, although it was a long journey.</b>	<b>Safely To meet similarity target</b>
			<b>Though it took many hours travel, we finally reached our house safely.</b>	<b>Paraphrased sentence. (SP88) Statement</b>

Medium-High 5

<b>Large Larger</b>		<b>250</b>	<b>Peter was a very large youth, whose size intimidated most people, much to his delight.</b>	<b>Adjective-Adjective Comparative</b>
	<b>M- H</b>	<b>789</b>	<b>Now I wouldn't say he was fat, but I'd certainly say he was one of the larger boys.</b>	
<b>Dry</b>		<b>672</b>	<b>We got soaked in the rain today, but now we are</b>	<b>Adjective Antonyms</b>

<b>Wet</b>			<b>nice and dry.</b>	
	<b>M-H</b>	<b>334</b>	<b>I was absolutely soaking wet last night, I drove my bike through the worst weather.</b>	
<b>Eventually Safely</b>		<b>522</b>	<b>After hours of getting lost we eventually arrived at the hotel.</b>	<b>High Frequency Adverbs</b>
	<b>M-H</b>	<b>39</b>	<b>After walking against the strong wind for hours he finally returned home safely.</b>	
<b>Difficult Mostly</b>		<b>396</b>	<b>It is mostly very difficult to gain full marks in today's exam.</b>	<b>To meet similarity target Adjective / Adverb Same word</b>
	<b>M-H</b>	<b>1017</b>	<b>The exam was really difficult, I've got no idea if I'm going to pass.</b>	
<b>Console</b>		<b>1117</b>	<b>If you don't console with a friend, there is a chance you may hurt their feelings.</b>	<b>To meet similarity target Verb Levin class 3.1 (Relax) Same word / same theme Good Friend</b>
	<b>M-H</b>	<b>605</b>	<b>One of the qualities of a good friend is the ability to console.</b>	

Medium - 10

<b>Far Farther</b>		<b>395</b>	<b>Will I have to drive far to get to the nearest petrol station?</b>	<b>Adverb Place Comparative Paraphrased sentence derived from far but using Farther instead. (SP112) Question</b>
			<b>Is it much farther for me to drive to the next gas station?</b>	
<b>Cat Hen</b>		<b>7</b>	<b>The cat was hungry so he went into the back garden to find lunch.</b>	<b>Living Biological Nouns</b>
	<b>M</b>	<b>1043</b>	<b>The hen walked about in the yard eating tasty grain.</b>	
<b>Pleasure Doubt</b>		<b>144</b>	<b>The pleasure that I get from studying, is that I learn new things.</b>	<b>Abstract emotion + vs –</b>
	<b>M</b>	<b>976</b>	<b>I have a doubt about this exam, we never got to study for it.</b>	
<b>Mostly Partially</b>		<b>600</b>	<b>The damp was mostly in the very corner of the room.</b>	<b>Two Low Frequency Adverbs Frequency / degree</b>
	<b>M</b>	<b>1072</b>	<b>The young lady was somewhat partially burnt from the sun.</b>	
<b>Drink Eat</b>		<b>389</b>	<b>Would you like to go out to drink with me tonight?</b>	<b>Levin Class 39.1</b>
	<b>M</b>	<b>86</b>	<b>I really don't know what to eat tonight so I might go out somewhere.</b>	
<b>Cover</b>		<b>631</b>	<b>You shouldn't be covering what you really feel.</b>	<b>State verb</b>
	<b>M</b>	<b>950</b>	<b>There is no point in covering up what you said, we all know.</b>	
<b>Appear</b>		<b>434</b>	<b>The ghost appeared from nowhere and frightened the old man.</b>	<b>Process verb</b>

	M	1066	The ghost of Queen Victoria appears to me every night, I don't know why, I don't even like the royals.	
Young		1098	If you continuously use these products, I guarantee you will look very young.	Adjective Commitment Paraphrased sentence (SP81)
			I assure you that, by using these products consistently over a long period of time, you will appear really young.	
Temperature		1050	Water freezes at a certain temperature, which is zero degrees Celsius.	Noun to allow word to participate
	M	210	The temperature of boiling water is 100 C and the temperature of ice is 0 C.	
Responsible		118	Because I am the eldest one I should be more responsible.	Noun To meet similarity target Contains a human error - 866 is not a question without prosodic data
	M	866	Just because of my age, people shouldn't think I'm a responsible adult, but they do?	

Medium-Low - 5

Nation		937	I am proud of our nation, well, most of it.	Abstract Noun
	M-L	609	I think of myself as being part of a nation.	
Fire		68	I need to dash into the kitchen because I think my chip pan is on fire.	To meet similarity target Noun Same word
	M-L	495	In the event of a chip pan fire follow the instructions on the safety note.	
Grey		217	I'm going to buy a grey jumper today, in half an hour.	To meet similarity target Adjective Same word
	M-L	670	That's a nice grey top, where did you get it from?	
Delay		288	When I was going out to meet my friends there was a delay at	Polysemous Noun/Verb



			<b>the train station.</b>	<b>combination</b>
	<b>M-L</b>	<b>1068</b>	<b>The train operator announced to the passengers that the train would be delayed.</b>	
<b>Mostly Most seriously</b>		<b>639</b>	<b>I'm worried most seriously about the presentation, not the essay.</b>	<b>Adverbs HF / LF and near homograph</b>
	<b>L-M</b>	<b>396</b>	<b>It is mostly very difficult to gain full marks in today's exam.</b>	

Low-Medium - 5

<b>Great Terrible</b>		<b>706</b>	<b>That's not a very good car, on the other hand mine is great.</b>	<b>Adjective Value + / -</b>
	<b>L-M</b>	<b>186</b>	<b>This is a terrible noise level for a new car.</b>	
<b>Seriously Difficult</b>		<b>30</b>	<b>I hope you're taking this seriously, if not you can get out of here.</b>	<b>Adverb Manner / Adjective Difficulty</b>
	<b>L-M</b>	<b>748</b>	<b>The difficult course meant that only the strong would survive.</b>	
<b>Conflict Prejudice</b>		<b>93</b>	<b>He fought in the war in Iraq before being killed in a car crash.</b>	<b>Abstract State / Emotion –ve affect</b>
	<b>L-M</b>	<b>780</b>	<b>The prejudice I suffered whilst on holiday in Iraq was quite alarming.</b>	
<b>Lemon Rose</b>		<b>142</b>	<b>My bedroom wall is lemon coloured but my mother says it is yellow.</b>	<b>Living Biological Fruits&amp;Vegs vs Plants&amp;Flowers</b>
	<b>L-M</b>	<b>325</b>	<b>Roses can be different colours, it has to be said red is the best though.</b>	
<b>Rotten Global</b>		<b>430</b>	<b>The perpetrators of war crimes are rotten to the core.</b>	<b>Two low frequency Adjectives</b>
	<b>L-M</b>	<b>158</b>	<b>There are many global issues that everybody should be aware of, such as the threat of terrorism.</b>	

Low - 15

<b>Guitar Weapon</b>		<b>657</b>	<b>I bought a new guitar today, do you like it?</b>	<b>Nouns Miscellaneous Artefact / Musical Instrument</b>
	<b>L</b>	<b>537</b>	<b>The weapon choice reflects the personality of the carrier.</b>	
<b>Relax Dry</b>		<b>1030</b>	<b>Does music help you to relax, or does it distract you too much?</b>	<b>Two State Process Verbs (human error – Dry adjective)</b>
	<b>L</b>	<b>1102</b>	<b>Does this sponge look wet or dry to you?</b>	
<b>Key Quay</b>		<b>2</b>	<b>This key doesn't seem to be working, could you give me another?</b>	<b>Homophone pair</b>
	<b>L</b>	<b>860</b>	<b>I dislike the word quay, it confuses me, I always think of things for locks, there's another one.</b>	
<b>Boat Chair</b>		<b>150</b>	<b>Boats come in all shapes and sizes but they all do the same thing.</b>	<b>Normally found indoors/outdoors</b>
	<b>L</b>	<b>322</b>	<b>Chairs can be comfy and not comfy, depending on the chair.</b>	
<b>Heap crowd</b>		<b>12</b>	<b>There was a heap of rubble left by the builders outside my house this morning.</b>	<b>Collective Nouns living/non-living</b>
	<b>VL</b>	<b>106</b>	<b>Sometimes in a large crowd accidents may happen, which can cause deadly injuries.</b>	
<b>Run Climb</b>		<b>338</b>	<b>I used to run quite a lot, in fact once I ran for North Tyneside.</b>	<b>Two Action Verbs Levin Class 51.3.2</b>
	<b>L</b>	<b>554</b>	<b>I used to climb lots at school as we had a new climbing wall put in the gym.</b>	
<b>Laugh Bargain</b>		<b>553</b>	<b>I love to laugh as it makes me happy as well as those around me.</b>	<b>Two Action Verbs, one LF</b>
	<b>VL</b>	<b>198</b>	<b>I thought we bargained that it would only cost me a pound.</b>	
<b>Punish</b>		<b>474</b>	<b>He was harshly punished for setting the fire alarms</b>	<b>Two Action-Process verbs</b>

Please turn over

Delay			off.	
	L	794	He delayed his response, in order to create a tense atmosphere.	
Change Appear		301	Someone spilt a drink accidentally on my shirt, so I changed it.	Two Process Verbs
	VL	675	It appears to have shrunk, it wasn't that size before I washed it.	
Safely Seriously		1073	The children crossed the road very safely thanks to the help of the lollipop lady.	Two Manner Adverbs
	VL	266	It was feared that the child might not recover, because he was seriously ill.	
Responsible Slow		1060	The responsible man felt very guilty when he crashed into the back of someone's car.	Two adjectives, to allow these words to participate
	L		A slow driver can be annoying even though they are driving safely.	
Fly Dry (verb)		321	Flies can also carry a lot of disease and cause maggots.	Noun-Verb, to allow these words to participate
	L	229	I dry my hair after I wash it or I will get ill.	
Hill		938	Meet me on the hill behind the church in half an hour.	Noun Hill Instruction Paraphrased sentence. (SP107)
			Join me on the hill at the back of the church in thirty minutes time.	
Hope cover		258	They said they were hoping to go to America on holiday.	Hope-cover Two state verbs
	L	340	I like to cover myself up in lots of layers, I don't like the cold.	
Drink		477	Would you like to drink this wine with your meal?	Drink Question Paraphrased sentence (SP120)
			Will you drink a glass of wine while you eat?	

Replaced speech acts

<b>pleasure</b>		<b>456</b>	<b>It gives me great pleasure to announce the winner of this year's beauty pageant.</b>	<b>Declaration Paraphrased sentence (SP110)</b>
			<b>It's a real pleasure to tell you who has won our annual beauty parade.</b>	
<b>heap</b>		<b>50</b>	<b>I am sorry but I can't go out as I have a heap of work to do.</b>	<b>Expression (SP114)</b>
			<b>I've a heap of things to finish so I can't go out I'm afraid.</b>	
<b>wet</b>		<b>189</b>	<b>Get that wet dog off my brand new white sofa.</b>	<b>Instruction (SP116)</b>
			<b>Make that wet hound get off my white couch – I only just bought it.</b>	
<b>climb</b>		<b>24</b>	<b>Could you climb up the tree and save my cat from jumping please?</b>	<b>Question (SP123)</b>
			<b>Can you get up that tree and rescue my cat otherwise it might jump?</b>	

The last 4 sentence pairs used 4 slots originally reserved for a factor-based study of the influence of dialogue acts. Further analysis suggested that more slots than were available were required for this, so the slots were re-allocated to balancing the similarity distribution.

## **I. CALL FOR PARTICIPANTS IN A STUDY OF SEMANTIC SIMILARITY**

### **Call for participants in a study of semantic similarity**

We would like to request your participation in a scientific study of semantic similarity. For ethical reasons we are required to ask your permission in advance and let you know what you are agreeing to. We have provided the answers to the key ethical questions below. If you require any further information before agreeing to participate please contact Jim O'Shea ([j.d.oshea@mmu.ac.uk](mailto:j.d.oshea@mmu.ac.uk), 0161 247 1546)

### **What will you ask me to do?**

If you agree, you will be asked to complete a form by writing a set of sentences. You will be supplied with a word to be used in the sentence and (in one case) a topic for the sentence.

You will also be asked to complete a few questions about yourself. These are your name, approximate age and the subject of the degree you are studying. You will also be asked to confirm that you are a native speaker of English (i.e. someone for whom it is their first language, spoken since birth).

The reason for this is that sometimes scientific studies produce surprising results, which need to be analysed, and this background information could help.

### **What will you expect me to know?**

We want to emphasise that we are not testing your intelligence or ability in any way, however we do expect you to know how to use Nouns, Adjectives, Verbs and Adverbs in a sentence.

### **Is there any risk?**

The words used to generate the sentences are ordinary English words that would not normally be considered offensive. The topics are taken from elementary language courses. The risk involved is equivalent to looking up an ordinary word in a dictionary.

### **How long will the data be kept for?**

The answers to the questions about yourself will be kept for no longer than 3 months after the first results are published.

The sentences you provide will be separated from the personal data and kept permanently. This is because the data can be very useful in long-term studies. Data of this type collected in the 1960s is still used widely today.

### **Will you publish my personal information?**

We will never disclose your personal information to anyone outside the project. Selected sentences from the set will be used in publications on an international scale. The complete set of sentences, made anonymous by removing personal details, will be made available to bona fide researchers on request.

### **Will I be compensated for participation?**

You will be compensated for contributing your time to help in this study with a payment of £5 on completion.

### **How long will it take?**

The task is timed to take a maximum of 1 hour, people who finish early will be asked to generate some extra sentences

## J. SENTENCE CAPTURE EXPERIMENT INSTRUCTION SHEET

### Thank-you for volunteering to take part in this study

You may still withdraw before starting the questionnaire or at any point before completing it, but the participation payment will only be made to people who complete the questionnaire.

Please start at the first page and work through the pages in order from start to end.

In this experiment we would like you to help us by writing some sentences each of which contains a particular word that we will supply.

Above all, we are looking for natural sentences. We would like you to write the kind of sentence that you would be likely to say to someone in a real conversation. If you can't do this try to think of a sentence that he would naturally write in an Internet forum, an e-mail or a letter. Alternatively, try to think of a sentence that someone might say to you in one of these circumstances. If all else fails try to think of a sentence that someone would say on a radio or television programme that you would listen to.

We would like you to think carefully about each sentence and please don't just jot down some cliché or proverb. Please don't stick to a single form of sentence; they can be statements, instructions, commitments, expressions and declarations.

Please note that the study does not evaluate you in any way, what we are testing in this experiment are the properties of natural English sentences. So there are no "right" or "wrong" answers, except in the sense that the right answer to each question is one that is natural for you.

Please also be aware that we will ask you to write **two sentences between 10 and 20 words long** in BLOCK CAPITALS for each word that we are interested in.

If you have any problems, questions or comments please speak to one of the investigators

. . . Otherwise please turn over and begin.

**Statement:**

Statements, descriptions, classifications, explanations

e.g.

Siamese cats are very rare in the part of town that I live in.

**Instruction:**

Instructions, orders, commands, requests

e.g.

Turn off all off the computers and the printer before you leave the lab tonight.

**Commitments:**

Promises, vows, pledges, contracts, guarantees

e.g.

I will pay you everything I owe by the end of the month.

**Expressions**

Apologies, thanks, congratulations, welcomes, condolences

e.g.

I am sorry about damaging your car in the car park last week.

**Declarations**

Declarations, pronouncements

e.g.

After counting the votes, Adam Taylor is the duly elected representative for Software Engineering.

## K. EXAMPLE EXTRACTS FROM QUESTIONNAIRES

Section introduction:

The next 3 pages are about VERBS

A verb, such as *to fight*, is a word that expresses an action or a state of being.  
You can use any valid form of the Verb in the sentence you write for example:

fight ... as in ... I, you, we, they ... fight  
fights ... as in ... he, she ... fights  
... fought ...  
... fighting ...

but NOT the Noun sense of the word, for example:

... the fight ...  
... a vicious fight ...  
... many fights ...

Please ask an investigator if you have any questions before continuing ...

otherwise please turn over



Example extracts from questionnaires normal sentence entry page

Please print (in BLOCK CAPITALS) **two** sentences, **between 10 and 20 words long** in the boxes below, using the Verb APPEAR

Sentence 1

Sentence 2

Please turn over

Example extracts from questionnaires themed sentence entry page

Please print (in BLOCK CAPITALS) **two** sentences, **between 10 and 20 words long** in the boxes below, on the general Topic of

**1.1 Going out (socially), giving invitations**

**and** using the Noun QUAY

Sentence 1

Sentence 2

Please turn over

**PARTICIPANT DETAILS (Non-student)**

And now, just a few items of personal information:

Your name (print)	
Your age	
Your highest educational qualification (including subject)	
Confirmation that you are a native English speaker* - please tick	<input type="checkbox"/>

Signature:

\* Native English speaker means that it is your first language and you have spoken it since birth.

**L. PARAPHRASE SENTENCE SUBSET (selected to generate additional very high similarity sentence pairs)**

Stim Word	Sentence	
Laugh Commit	I will make you laugh so much that your sides ache.	71
Parent Express	I offer my condolences to the parents of John Smith, who was unfortunately murdered.	79
Young Commit	If you continuously use these products, I guarantee you will look very young.	81
Safely State	We got home safely in the end, although it was a long journey.	88
Far/farther Quest	Will I have to drive far to get to the nearest petrol station?	112
Hill Instruct	Meet me on the hill behind the church in half an hour.	107
Drink Quest	Would you like to drink this wine with your meal?	120
Pleasure Declare	It gives me great pleasure to announce the winner of this year's beauty pageant.	110
Heap Express	I am sorry but I can't go out as I have a heap of work to do.	114
Wet Instruct	Get that wet dog off my brand new white sofa.	116
Climb Quest	Could you climb up the tree and save my cat from jumping please?	123

Type of Dialogue Act is indicated below the stimulus word

## M. SAMPLE PARAPHRASE CAPTURE SHEET

Thank-you for agreeing to help with this study. You can withdraw from the study if you wish at any time before returning this questionnaire.

This study is not testing you in any way, it is to produce data that can be used in future experiments which measure the similarity of sentence pairs.

We would like you to paraphrase 3 sentences. By paraphrase we mean we want you to express the same sentence in a different way.

In each case the sentence you write should come as close as you can manage to meaning the same thing as the example supplied to you

The sentences should be between 10 and 20 words long

You will be told that there is one word that you **MUST** use in the sentence you write

Otherwise please try not to use more words form the original sentence than you have to

Please **PRINT** the sentences in the boxes provided as they will have to be typed into a computer

1. Your sentence must contain the verb **LAUGH**

Example	I will make you laugh so much that your sides ache.
Your version	

2. Your sentence must contain the adverb **SAFELY**

Example	We got home safely in the end, although it was a long journey.
Your version	

3. Your sentence must contain the noun **PARENTS**

Example	I offer my condolences to the parents of John Smith, who was unfortunately murdered.
Your version	

## **N. CALL FOR PARTICIPANTS IN AN EXPERIMENTAL STUDY OF SENTENCE SEMANTIC SIMILARITY.**

We would like to request your participation in a scientific study of semantic similarity. For ethical reasons we are required to ask your permission in advance and let you know what you are agreeing to. We have provided the answers to the key ethical questions below. If you require any further information before agreeing to participate please contact Jim O'Shea ([j.d.oshea@mmu.ac.uk](mailto:j.d.oshea@mmu.ac.uk), 0161 247 1546)

### **What will you ask me to do?**

If you agree you will be asked to sort a pack of 66 cards containing pairs of sentences. Then you will be asked to write down a rating for the similarity of meaning of each pair of sentences.

You will also be asked to complete a few questions about yourself.

These are your name, age and highest level of qualification. You will also be asked to confirm that you are a native speaker of English (i.e. someone for whom it is their first language, spoken since birth).

We ask for some personal data because sometimes scientific studies produce surprising results which need to be analysed and this background information could help.

### **Is there any risk?**

The sentences do not contain any words which would generally be considered to be offensive. The risk involved is equivalent to looking up an ordinary word in a dictionary.

### **How long will the data be kept for?**

The answers to the questions about yourself will be kept for no longer than is necessary to check for errors or interesting properties of the data. This will be for no longer than 3 months after the first results are published.

The ratings you provide will be separated from the personal data and kept permanently. This is because the data can be very useful in long-term studies. Data of this type collected in the 1960s is still used widely today.

### **Will you publish my personal information?**

We will never disclose your personal information to anyone outside the project. Statistical summaries of the ratings will be published on an international scale. The set of individual ratings, made anonymous by removing personal details, may be made available to bona fide researchers on request.

### **Will I be compensated for participation?**

You will be compensated for contributing your time to help in this study with a payment of £5 on completion.

### **How long will it take?**

The task is timed to take a maximum of 1 hour.

## O. SURVEY: SENTENCE SEMANTIC SIMILARITY

### **Background information – please read before you start doing the task**

Thank-you for volunteering to take part in this study.

You may still withdraw before starting the task or at any point while doing it.

You are provided with an envelope containing a set of cards and a recording sheet to write your judgements on (please don't write anything on the cards). The cards have been shuffled into a random order.

Each card has two sentences written on it. We want you to rate the similarity of meaning of these sentence pairs.

What do we mean by similarity of meaning?

To judge similarity of meaning you should look at the two sentences and ask yourself "How close do these two sentences come to meaning the same thing?"

In other words:

How close do they come to making you believe the same thing?

How close do they come to making you feel the same thing?

or

How close do they come to making you do the same thing?

You will be asked to sort the cards into 4 piles, in order of similarity.

- The highest similarity pile is for sentence pairs which mean exactly the same thing, plus others with a high similarity.
- The lowest similarity pile is for sentence pairs whose meanings have no connection whatsoever, plus others with a low similarity.
- The other two piles will contain sentences that fall somewhere between the lowest and highest similarities.

We don't have any expectations about how many sentences will be in each pile, again this entirely up to you to decide.

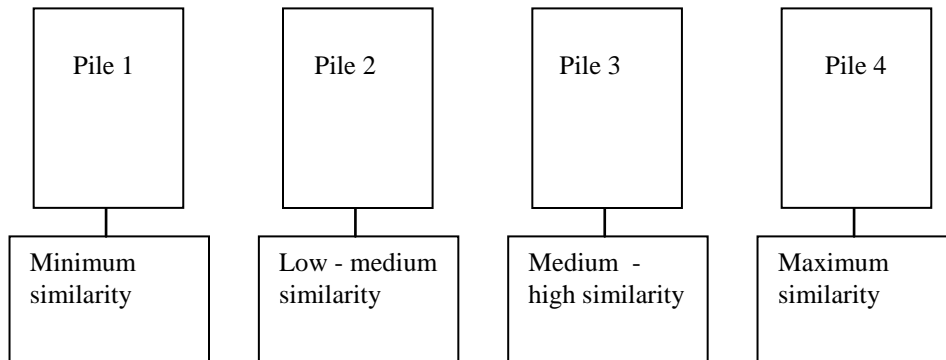
Some of the sentences occur more than once and sometimes they have small changes in the wording, so please read each sentence pair carefully before making your judgement.

If you have any problems, questions or comments please speak to one of the investigators.

Please note that the study does not evaluate you in any way – there are no "right" or "wrong" answers, except in the sense that the right answer to each question is an accurate expression of your personal opinion.

## Instructions

(1) Please open your envelope and sort the cards that you find inside into 4 piles, in ascending order of similarity of meaning.



(2) Now please read the cards in each pile carefully. If you change your opinion of which pile it should be in, please move it to the other pile.

(3) Finally, please rate the similarity of each pair of sentences by writing a number between 0.0 (minimum similarity) and 4.0 (maximum similarity) on the recording sheet. Please do not use values greater than 4.0. You can use the first decimal place (e.g. 2.2) to show finer degrees of similarity.

If you have any problems, questions or comments please speak to one of the investigators.

### Guidance Notes

If you have difficulty making an assessment here are some descriptions of the main points on the scale to help you:

- 0.0 The sentences are unrelated in meaning.
- 1.0 The sentences are vaguely similar in meaning.
- 2.0 The sentences are very much alike in meaning.
- 3.0 The sentences are strongly related in meaning.
- 4.0 The sentences are identical in meaning.

You can use the first decimal place, for example if you think the similarity is half way between 3.0 and 4.0 you can use a value like 3.5.

Please turn over



SAMPLE CARDS WITH SENTENCE PAIRS

Sentence Pair 76

**You're not a good friend if you're not prepared to be present when I need you.**

**A good friend always seems to be present when you need them.**

Please follow the instructions about sorting these cards before writing down your ratings

Sentence Pair 122

**Flies can also carry a lot of disease and cause maggots.**

**I dry my hair after I wash it or I will get ill.**

Please follow the instructions about sorting these cards before writing down your ratings

Sentence Pair 121

**Roses can be different colours, it has to be said red is the best though.**

**Roses come in many varieties and colours , but yellow is my favourite.**

Please follow the instructions about sorting these cards before writing down your ratings

Sentence Pair 128

**I hope you're taking this seriously, if not you can get out of here.**

**The difficult course meant that only the strong would survive.**

Please follow the instructions about sorting these cards before writing down your ratings

**SIMILARITY RATING SHEET (sheet 1 of 2)**

Please enter a rating for the similarity of meaning of each sentence pair.

The rating scale runs from 0.0 (minimum similarity) to 4.0 (maximum similarity), please do not use values greater than 4.0.

SP66	
SP67	
SP68	
SP69	
SP70	
SP71	
SP72	
SP73	
SP74	
SP75	
SP76	
SP77	
SP78	
SP79	

SP80	
SP81	
SP82	
SP83	
SP84	
SP85	
SP86	
SP87	
SP88	
SP89	
SP90	
SP91	
SP92	
SP93	
SP94	

SP95	
SP96	
SP97	
SP98	
SP99	
SP100	
SP101	
SP102	
SP103	
SP104	
SP105	
SP106	
SP107	
SP108	
SP109	

**Guidance Notes**

If you have difficulty making an assessment here are some descriptions of the main points on the scale to help you:

- 0.0 The sentences are unrelated in meaning.
- 1.0 The sentences are vaguely similar in meaning.
- 2.0 The sentences are very much alike in meaning.
- 3.0 The sentences are strongly related in meaning.
- 4.0 The sentences are identical in meaning.

You can use the first decimal place, for example if you think the similarity is half way between 3.0 and 4.0 you can use a value like 3.5.

**PARTICIPANT DETAILS (Student)**

And finally, a few details about yourself . . .

. These are your name, approximate age and the subject of the degree you are studying. You will also be asked to confirm that you are a native speaker of English (i.e. someone for whom it is their first language, spoken since birth).

Name			
Age (tick)	Under 18 <input type="checkbox"/>	18-22 <input type="checkbox"/>	Older than 22 <input type="checkbox"/>
Degree Title			
I confirm that I am a Native English speaker (it is my first language, spoken since birth)	Sign below:		

Please collect your compensation and sign a receipt.

## REFERENCES

- O'SHEA, J. D., et al. 2008. *A comparative study of two short text semantic similarity measures*. Lecture Notes in Artificial Intelligence 4953/2008, 172-181.
- O'SHEA, J. 2010. A framework for applying short text semantic similarity in goal-oriented conversational agents. Manchester Metropolitan University.
- MILLER, G. A. and CHARLES, W. G. 1991. *Contextual correlates of semantic similarity*. Language and Cognitive Processes 6 1-28.
- UITENBROEK, D. G., 2013. Simple statistical correlation analysis online. <http://www.quantitativeskills.com/sisa/statistics/correl.htm>
- GRABIN, C., 2013. General statistics. <http://psych.unl.edu/psycrs/statpage/regression.html>

## BIBLIOGRAPHY

- LI, Y., BANDAR, Z., McLEAN, D. & O'SHEA, J. 2006 Sentence Similarity Based on Semantic Nets and Corpus Statistics. *IEEE Trans. Knowl. Data Eng.*, 18, 1138-1150.