CORPUS ANALYSIS
Marta Puliga

Extended Abstracts / Summaries of Articles Reporting Studies about Depression in Psychology (in order to be published on the Internet or in a journal)

Total number of ***Words***	5704
Total number of ***Word Types***	1442

QUANTITATIVE ANALYSIS

1) The Most Frequent 100 Words Used in the Corpus

[1]	the	294
[2]	and	199
[3]	of	189
[4]	to	137
[5]	in	114
[6]	a	108
[7]	that	76
[8]	with	53
[9]	depression	45
[10]	for	44
[11]	s	43
[12]	as	42
[13]	is	41
[14]	more	41
[15]	their	38
[16]	were	38
[17]	by	37
[18]	at	36
[19]	have	34
[20]	or	32
[21]	negative	31
[22]	health	30
[23]	study	30
[24]	adults	29
[25]	may	29
[26]	people	29
[27]	they	29
[28]	are	28
[29]	be	28
[30]	it	28
[31]	not	28
[32]	who	28
[33]	from	26
[34]	on	26
[35]	this	26
[36]	depressed	25
[37]	self	25

[38]	had	24
[39]	other	23
[40]	than	22
[41]	these	22
[42]	but	20
[43]	one	19
[44]	research	19
[45]	researchers	19
[46]	age	18
[47]	found	18
[48]	also	17
[49]	authors	17
[50]	children	17
[51]	images	17
[52]	life	17
[53]	problems	17
[54]	was	17
[55]	an	16
[56]	d	16
[57]	esteem	16
[58]	how	16
[59]	mental	16
[60]	findings	15
[61]	mothers	15
[62]	older	15
[63]	participants	15
[64]	two	15
[65]	which	15
[66]	anxiety	14
[67]	during	14
[68]	high	14
[69]	percent	14
[70]	psychological	14
[71]	about	13
[72]	after	13
[73]	association	13
[74]	can	13

[75]	disorder	13
[76]	experienced	13
[77]	functioning	13
[78]	mdd	13
[79]	physical	13
[80]	positive	13
[81]	those	13
[82]	when	13
[83]	crh	12
[84]	disorders	12
[85]	ph	12
[86]	said	12
[87]	such	12
[88]	support	12
[89]	university	12
[90]	according	11
[91]	being	11
[92]	between	11
[93]	each	11
[94]	his	11
[95]	person	11
[96]	says	11
[97]	showed	11
[98]	what	11
[99]	adolescence	10
[100]	been	10
[101]	brain	10
[102]	colleagues	10

There are 102 items instead of 100 because the items number 56 and 85 are part of a personal noun, and so we cannot consider them.

2) The Most Significative Items in the List

Nouns	30 items

1	[9]	depression	45
2	[23]	study	36
3	[22]	health	30
4	[24]	adults	38
5	[26]	people	29
6	[37]	self-	25
7	[75]	disorder	25
8	[46]	age	22
9	[44]	research	19
10	[45]	researchers	19
11	[49]	authors	17
12	[50]	children	17
13	[51]	images	17
14	[52]	life	17
15	[53]	problems	17
16	[57]	esteem	16
17	[60]	findings	15
18	[61]	mothers	15
19	[63]	participants	15
20	[66]	anxiety	14
21	[69]	percent	14
22	[73]	association	13
23	[161]	group*	13
24	[77]	functioning	13
25	[88]	support	12
26	[89]	university	12
27	[95]	person	11
28	[99]	adolescence	10
29	[101]	brain	10
30	[102]	colleagues	10

*** The occurrences of the word “group” are the result of the addition between the occurrences of the noun in the singular form (7) and in the plural form (6).**

All Verbs	20 items	What will be considered are the verbs which appear in the first 200 words of the frequency list, since there are some verbs worth noting

1	[13]	is	41
2	[16]	were	38
3	[19]	have	34
4	[25]	may	29
5	[28]	are	28
6	[29]	be	28
7	[38]	had	24
8	[47]	found	18
9	[54]	was	17
10	[74]	can	13
11	[76]	experienced	13
12	[86]	said	12
13	[91]	being	7 out of 11
14	[96]	says	11
15	[97]	showed	11
16	[100]	been	10
17	[148]	say	8
18	[162]	has	7
19	[171]	suggest/s	7+2
20	[175]	appears	6

Auxiliaries

[13]	is	41
[16]	were	38
[19]	have	34
[28]	are	28
[29]	be	28
[38]	had	24
[54]	was	17

Modals

[25]	may	29
[74]	can	13

Present tenses

Simple Present		Present Continuous	-Ing form as infinite	
is	40 out of 41	1 out of 41	Being	7 out of 11 *(4 times as a noun)*
have	19 out of 34			
may	29			
are	28			
can	13			
says/say	11+8			
has	2 out of 7			
suggest/s	7+2			
appears	6			

Past tenses

Simple Past		Perfect Tenses		Past Participle	
were	34 out of 38 *(4 times in past continuous)*				
		Have + pp	15 out of 34		
had	8 out of 24	Had + pp	14 out of 24	2 out of 24	
		Has + pp	5 out of 7		
found	12 out of 18			6 out of 18	
was	17				
experienced	8 out of 13			5 out of 13	
said	12				
showed	11			been	10

Passive Forms

1	is influenced by	The form "is" appears 17 times in a passive construction out of 41 occurrences
2	is influenced not only by	
3	is associated with	
4	is Associated With	
5	is entirely explained by	
6	is made	
7	is published by	
8	is associated with	
9	is published by	
10	is activated	
11	is activated	
12	is reported on	
13	is related to	
14	is found in	
15	is needed to see	
16	is involved in	
17	is not known	

1	are reported on	The form "are" appears 8 times in a passive construction out of 29 occurrences
2	are not necessarily tied	
3	are likely to have	
4	are likely to have	
5	are reported on	
6	are associated with	
7	are released	
8	are likely to demonstrate	

1	were examined to	The form "were" appears 9 times in a passive construction out of 38 occurrences
2	were asked	
3	were also assessed	
4	were interviewed	
5	were questioned	
6	were linked to	
7	were re-employed	
8	were first assessed	
9	were controlled	

1	was associated with	The form "was" appears 7 times in a passive construction out of 17 occurrences
2	was predicted by	
3	was found to influence	
4	was associated with	
5	was associated with	
6	was eliminated	
7	was prepared by	

1	has long been known
2	have been found
3	who've been diagnosed

1	will be presented	The form "be" appears 8 times in a passive construction out of 28 occurrences
2	could be generalized	
3	can be predicted	
4	could be exposed to	
5	to be small or poorly developed	
6	should be seen	
7	should be watched	
8	should be taught to	

Tot. number of passive forms: 52

All Adjectives | 16 items

1	[15]	their	38
2	[21]	negative	31
3	[35]	this	26
4	[36]	depressed	25
5	[39]	other	23
6	[41]	these	22
7	[59]	mental	16
8	[62]	older	15
9	[64]	two	15
10	[70]	psychological	14
11	[79]	physical	13
12	[80]	positive	13
13	[81]	those	13
14	[93]	each	11
15	[94]	his	11
16	[103]	depressive	10

Different kinds of adjectives

Qualifying adj		Demonstrative adj		Indefinite adj		Possessive adj		Numeral adj	
negative	31	this	23 out of 26	other	18 out of 23	their	38	two	15
depressed	25	these	21 out of 22	each	8 out of 11	his	11		
mental	16	those	4 out of 13						
older *comparative*	15								
psychological	14								
physical	13								
positive	13								
depressive	10								

Pronouns | 11 items

Personal p		Demonstrative p		Indefinite p		Relative p	
they	29	this	3 out of 26	other	5 out of 23	that	23 out of 76
it	28	these	1 out of 22	each	3 out of 11	who	28

		those	7 out of 13			which	15
						what	11

Adverbs	0 items

3) Minimal Core Vocabulary

NOUNS
adolescence
adults
age
anxiety
association
authors
brain
children
colleagues
depression
disorder
esteem
findings
functioning
group*
health
images
life
mothers
participants
people
percent
person
problems
research
researchers
self-
study
support
university

ADJECTIVES
depressed
depressive
each
his
mental
negative
old
other
physical
positive
psychological
their
these
this
those
two

VERBS	Forms which appear in the first 100 words of the frequency list
To appear	appears
To be	is, are, were, was, been, being
Can	can
To experience	experienced
To find	found
To have	have, has, had
May	may
To say	Say, says, said
To show	showed
To suggest	Suggest, suggests

QUALITATIVE ANALYSIS

1) Meaningful words: how they are used

Depression

It is the most frequent noun in the list, and this is justified by the fact that Depression is the topic of the abstracts / articles contained in the corpus. There are some features worth noting about this term:

a) It is often connected to the term **"anxiety"** (9 times), which very rarely appears far from this word in the texts. They are considered almost like a reversible binomial, which means that they appear both the following ways: *depression and anxiety* / *anxiety and depression*. It seems that Depression can cause anxiety and vice versa.
b) It is often connected to the term **"spouse"**(8 times), which usually appears connected to the binomial *anxiety and depression* analysed before: (e.g.) *spouse's level of anxiety and depression*.
c) It appears 8 times with the term **"level"** in the expression: *level of depression*.
d) In 5 occurrences, it is in connection with the word **"postnatal"**: *postnatal depression*.
e) It appears near the word **"adolescence"** 10 times.
f) It is also meaningful, but quite obvious, that there is a high frequency of the two adjectives derived from this noun: **depressed** (22 occurrences) and **depressive** (10 occurrences). The last one also appears in the acronym "MDD" (13 occurrences) which means Major Depressive **Disorder**, that is a kind of depression.

From these data, we can deduce that depression is considered as a *disorder* in psychology, and there can be different kinds of it, and each one can be at different levels, depending on the gravity. The different kinds of depression appearing in the corpus are: **spouse's depression, postnatal depression in mothers, depression in adolescents.**

Disorder

This is a very important technical term, which occurs 25 times as a single word (13 in the plural form and 12 in the singular form) and 13 times in the acronym MDD. There are different kinds of "disorders", one of which is, of course, depression. The kind of disorder is always expressed just placing the word defining the disorder before this term (premodification):

Examples

- ★ *Human brain disorders*
- ★ *Depressive disorder*
- ★ *Mental disorders*
- ★ *Post-traumatic stress disorder*
- ★ *Obsessive-compulsive disorder*

It is worth noting that **disorder** is the **psychological term** used to define what in the **everyday language** is defined a **mental illness.**

Study

This term occurs 30 times in the corpus, and it is used to talk about the research and the experiments at the basis of each article. It can be considered as a **synonym** of the term **research**, but they are used in a different way in the corpus: **study** is usually referred to what is the basis of the article, while **research** is usually referred to previous or possible future studies different from the one about which the article is built.

Health

There are 30 occurrences of this term. It often appears together with the adjectives **mental** (*mental health* 9 times) or **physical** (*physical health* 8 times). It usually appears before the words **problems, conditions, status**, (*health problems, health conditions, health status*) and usually referred to adult people.

May

May is the most frequent modal in the corpus. It occurs 29 times, 9 of which followed by **"be"**: *may be*. There is a high use of modals, if we also consider that **"can"** appears 13 times in the corpus. They are used to express a very high probability of a specific result that a particular condition can have, according to which what a study has stressed. This means that when showing the results of a psychological research, it is usual to never express absolute certainty, but always talk in terms of a big probability. Modals are one way of indicating the degree of certainty of a writer's commitment to a statement or claim; in other words, it is a way through which the writer distances from the statements. This phenomenon is generally known as *Hedging*.

Examples:

- ⋆ *life events such as job loss **can** have lasting effects*
- ⋆ *a depressed partner **may** have a negative worldview*
- ⋆ *adult health **may** be largely due to*

Self-

It always appears as a prefix, never as a suffix or a single noun. It occurs 16 times out of 25 together with the word **"esteem"**, in the compound noun: self-esteem.
This term **"self-esteem"** often appears connected with words like *happiness*, *happy*, *unhappy*, so that we can deduce that self-esteem can influence the happiness of a person a lot.

To experience

This verb occurs 13 times in the Past Simple and 4 times in the –ing form. It is worth noting that it is the term used to express the status or conditions a person has been in. In everyday language it is not so common to use it, since it is often replaced by "have".

Examples:

- ⋆ *those who **experienced** other psychological disorders* (everyday language: *those who **had**...*)
- ⋆ *a purpose in life and **experienced** good health* (everyday language: *...and **had** good health*)

2) Typical Linguistic Patterns and their Functions

To be associated with

It is a pattern which is always used to express the reversible cause-effect relationship between two conditions.

Examples:

- ⋆ depression in adolescence ***is associated with*** many subsequent problems
- ⋆ responses to frustration ***are associated with*** violent behavior
- ⋆ support during childhood ***is associated with*** increased levels of depression

Suggest/s that

It is the structure used to express the possible conclusions researchers can draw starting from the findings obtained in a study/research.

Examples:

- ⋆ The research ***suggests that***
- ⋆ our study ***suggests that***
- ⋆ These findings ***suggest that***

As we can see, the tone of the speech is *never in terms of absolute certainty* (see above "may"), but of *"suggestion"*. The use of this reporting verb is another way of expressing *hedging*.

Found that

It is the structure used to introduce the findings, results, of a research/study. It is worth noting that the synonym "discovered", accepted in everyday language, is never used in

the corpus, so that we can consider "found that" as a pattern typical of abstracts / articles reporting a study.

Examples:

- ⋆ Studies have also ***found that***
- ⋆ research groups have ***found that***
- ⋆ the researchers ***found that***

3) Nominalizations

Fighting, findings, functioning, learning, development, impairment, improvement, re-employment, treatment, unemployment, activation, association, correlation, invitation, presentations, ability, activity, hyperactivity, reliability, selectivity.

4) Compound nouns (nominal phrases)

Compound Nouns are used to express the different kinds of disorders/ illnesses/phobias. As written above (see the word "disorder") the compositional process is always the same: ***a noun defining the kind of disorder + disorder/phobias/….*** This process is referred to as **premodification** or **construction at left**, which means that anything that appears before a noun is to be considered a modifier.

Examples:

- ⋆ *disruptive behavior disorders*
- ⋆ *human anxiety disorders*
- ⋆ *panic disorder*
- ⋆ *stress disorder*
- ⋆ *adolescent non-mood disorder*
- ⋆ *attention deficit hyperactivity disorder*
- ⋆ *human brain disorders*

5) Passive Forms

Passive forms are quite frequent in the corpus, and this is a typical feature of Scientific English. They usually occurs in the third person singular of Simple Present (17 times out of 41 occurrences of the form **"is"** = 41.5%), but they are also frequent with the other forms of the verb to be in the corpus: **are** (third person plural) = 8 occurrences out of 29 (27.6%); **was** (third person singular, Past Simple) = 7 out of 17 (41%); **were** (third person plural, Past Simple) = 9 out of 38 (23.6%); **be** (base form appearing after a modal verb) = 8 out of 28 (28.6%).

As we can see, the use of passive forms emerges as a very important feature of the corpus, since in everyday language passive forms are not so common.

6) Structure of the summaries of articles reporting studies/researches on depression in psychology contained in the corpus

Each abstract/summary shows the same precise structure, so that we can consider it as the typical structure of these kind of summaries of articles reporting studies/researches in general. This is the repeated pattern followed in the corpus:

- ★ **Title**

- ★ **Introduction (very concise abstract)**
  - Name of the town where the study / research took place
  - A few lines resuming the objective of the research and the important findings obtained with it
  - References of the journal / book where the whole study is published

- ★ **Section 1**
  - Description of how the research was carried out: the examined groups of people: how, why, where

- ★ **Section 2**
  - Results and findings of the study / research

- ★ **Section 3**
  - Conclusions that can be drawn starting from the findings obtained in the experiment / research

Conclusions

Part of the conclusions I drew during the corpus analysis are included in the previous section (Qualitative Analysis).
Other worth noting conclusions:

- ★ There are very few logical connectors. The reason can be that these texts are summaries, which means that they have to be quite concise.
- ★ The most common verbal tense is **Present Simple**, which is typical of Scientific English, followed by **Simple Past**. There is also a **high frequency of the Passive Form** (if confronted with general everyday English,) which is another important feature of Scientific English. The **verbs are all in the third person (singular or plural)**, which can be explained by two important reasons: a) the necessity to show objectivity in the texts, keeping distances from what written, which can be seen as another example of *Hedging* ; b) these articles are summaries of very large studies, carried out by psychologists who are usually different from the authors of these summarizing articles.
- ★ **Personal pronouns** (third person singular and plural) and relative pronouns are quite frequent, such as **demonstrative** (apart from "that") and **possessive** (third person singular and plural) **adjectives**.
- ★ As shown in the qualitative analysis, apart from the auxiliaries and modals, the **most frequent verbs** (*to find, to suggest, to say, to show, to experience*) deal with the particular area of "reporting research and experiments", which is what these article do. As for the **most frequent nouns and qualifying adjectives** (*depression, study, health, disorder (...), negative, depressed, mental (...),*) they all deal with the topic of the articles, ("studies/researches/experiment on depression in psychology"), so that we can say they are **part of the technical vocabulary**, with a very specific meaning, according to the psychological field.
- ★ There is a very high frequency of **"that"** with a **declarative/explicative function**, which is also one of the main functions present in the corpus, as we can see in the following expressions: ***found that, suggest that.*** These are very frequent expressions, which can be considered as pattern-expressions typical of these kind of articles.
- ★ Another important frequent function in the corpus is the **expression of the cause-effect relationship**, mainly expressed through the following pattern: ***to be associated with***.
- ★ **Nominalizations and nominal phrases** are quite frequent (see points 3-4 in the previous section.) This is quite a common feature in Scientific English, since it enables complex information to be packaged into a phrase that is simple from a grammatical point of view.
- ★ As for the use of **Modals**, see point 1 (*May*) in the previous section (Quantitative Analysis.)

Marta Puliga