

Creation of a Corpus of Specialised Language

Simona Mosca

Checklist

QUANTITATIVE ANALYSIS

The corpus is based on material found on the Web, concerning **Medicine**:
hypertension and blood pressure measurement.

1. What are the most significant items in that list?

Frequency list of adjectives

Qualifier	Frequency
high	150
systolic	85
ambulatory	64
diastolic	58
cardiovascular	56
normal	39
important	32
clinic	30
white	23
Demonstrative	Frequency
that	146
this	78
those	26
these	43
Possessive	Frequency
your	62
their	32

Frequency list of adverbs

Adverbs	Frequency
as	123
than	80
more	76
over	46
also	45

when	44
not	40
such	31
there	28
however	27
because	26
only	24

Frequency list of conjunctions

Conjunctions	Frequency
and	453
or	182
for	173
as	123
by	80
than	80
from	60
but	50
if	45
when	44
not	40
however	27
because	26

Frequency list of nouns

Nouns	Frequency
blood	542
pressure	538
hypertension	217
risk	163
heart	95
patients	74
measurement	66
people	59
treatment	53
disease	47
one	46
stroke	38
factors	35

measurements	33
use	32
evidence	29
women	29
cause	28
number	28
values	28
damage	26
devices	26
home	26
levels	26
patient	26
studies	26
years	26
diabetes	25
time	25
kidney	24
coat	23
men	23
symptoms	23

Frequency list of prepositions

Prepositions	Frequency
of	557
in	438
to	349
for	173
with	162
as	123
on	103
by	80
than	80
from	60
but	50
over	46
between	31
during	28

Frequency list of pronouns

Dimostrativi	Frequency
---------------------	------------------

that	146
this	78
these	43
those	26
Personal	Frequency
it	73
you	34
Impersonal	Frequency
there	28
Relative	Frequency
which	49
when	44
who	31
Indefinite	Frequency
some	42
other	26
both	22

Frequency list of verbs

Auxiliar	Frequency
is	285
are	138
have	107
has	45
Infinitive	Frequency
be	129
Modal	Frequency
may	101
can	74
should	41
Past participle	Frequency
been	44
associated	26
called	25
used	24

Frequency lists of articles

Definite	Frequency
-----------------	------------------

the	759
Indefinite	Frequency
a	293
an	65

Core vocabulary:

Nouns	Adjectives	Verbs	Adverbs	Prepositions
adults	all	are	all	above
age	ambulatory	associated	also	as
alcohol	american	based	any	between
arm	any	be	as	but
artery	available	been	because	by
association	both	called	both	during
attack	cardiovascular	can	first	for
average	clinic	cause	however	from
blood	clinical	does	less	in
body	common	due	more	least
cases	conventional	elevated	most	less
cause	coronary	estimated	not	of
cholesterol	diastolic	has	only	on
coat	elderly	have	or	over
complications	essential	include	over	than
condition	even	including	so	there
control	excessive	increase	such	to
coronary	first	increased	than	with
cuff	greater	increases	there	
damage	high	increasing	usually	
data	higher	is	well	
day	important	isolated	when	
	daytime	increasing	may	
	devices	its	measured	
	diabetes	less	need	
	disease	low	provide	
	doctor	lower	raised	
	drugs	major	reduce	
	effect	medical	reduced	
	events	normal	reported	
	evidence	older	see	
	exercise	only	should	

fact	primary	shown
factor	recent	sleep
failure	secondary	stroke
guidelines	short	taken
health	systolic	used
heart	their	using
history	these	was
home	this	were
hypertension	those	will

increase	well
information	white
insulin	your

intake
interventions
kidney
level
lifestyle
major
management
measurement
measuring
men
mercury
monitoring
need
night
number
obesity
office
one
organ
patient
people
period
population
practice
predictor
pressure
pulse
rate
reading
reduction
results

risk
salt
signs
smoking
society
sodium
stress
stroke
studies
subjects
symptoms
system
table
target
technique
term
tests
thresholds
time
treatment
trials
two
use
values
vessels
weight
women
year

QUALITATIVE ANALYSIS

- 1. Focus on some of the most frequent words of your list, for example technical or semi-technical vocabulary, modals, verbs, connectors, etc.**
- 2. Start observing their collocation.**

3. Is there any linguistic pattern that is typical of the target language?

- **Hypertension** (In everyday language we say *high blood pressure*)
- **Estimated** (ex. About 40% of patients **were estimated** to achieve a reduction in systolic blood pressure...; Brief interventions **have been estimated** to cost between £40 and £60...; ...**estimated differences** between ambulatory and clinic values vary with these factors.; **An estimated 10% of hypertension cases** are caused by alcohol abuse ...). In everyday language we would say *considered*.
- **Shown** (ex. ... as **shown** in recent randomized controlled trials; ... has been **shown** in some observational studies to be a better predictor ...; Although this database has been **shown** to be reasonably applicable to ...). In everyday language we would say *demonstrated*.

4. What are the most common forms of nominalization (if any)?

- Measurement (66)
- Treatment (53)
- Stroke (38)
- Studies (26)
- Monitoring (22)
- Reading (20)
- Information (18)
- Complications (17)
- Management (15)
- Measuring (14)
- Smoking (14)
- Reduction (13)
- Population (11)
- Interventions (11)
- Association (11)

5. From the observation of certain items, can you detect the presence of specific language functions essential in the target language?

- **Blood pressure measurement / Blood pressure can be / should be / is measured ...** (ex. The blood pressure **is measured**...)
- **To be at risk** (ex. ...people with hypertension **are at higher risk** for heart attacks and stroke ...)
- **Treatment is based on** (ex. In one controlled trial⁶⁴ **treatment based on** ABPM...)
- **Intervention(s) is/are based on** (ex. **Intervention** might **be based on** relative risk...)

6. Are there any passive forms? How often are they used and when?

There is a large presence of **passive form** of the **perfect tenses** [**have been ...ed (15) / has been ...ed (18)**] and the **simple tenses** [**is/are ...ed or was ...ed (15)/were ...ed (15)**]. They are used to explain the results or observations about studies or testing.

Examples:

- Complex and computerized methods **have been developed** for estimating short-term risk.
- This database **has been shown** to be reasonably applicable to some European populations...
- Other signs and symptoms sometimes associated with high blood pressure generally **are caused** by other conditions...
- According to these guidelines, blood pressure for adults **is classified** as follows...
- Interventions actively combining exercise and diet **were shown** to reduce both systolic and diastolic blood pressure...
- In a recent study, the persistent excess of CHD events in treated hypertensive subjects **was predicted** by three factors...

7. What are the most common verbal tenses to be found?

- **Passive forms**, used to explain the results or observations on studies or testing. (ex. Excessive alcohol consumption **is associated** with raised blood pressure and poorer cardiovascular and hepatic health.)
- **Modal verbs and Imperative**, used to give instructions for blood pressure measurement (examples: In patients with symptoms of postural hypotension **measure** blood pressure while patient is standing...; Home blood pressure measurements for suitable periods before and during treatment **can** also be recommended because this relatively cheap procedure **may** improve patient's adherence to treatment...; Therefore, the diagnosis of hypertension **should** be based on multiple blood pressure measurements, taken on separate occasions).

8. Can you detect compound noun phrases that are typical of this variety, (i.e. nominal phrases)?

We can find, for example,

- **blood pressure measurement**
- **blood pressure reduction**

9. Some more information ...

In my corpus there is a big presence of **comparatives** and **the more... and the more...** form, used particularly to compare the different ways of measuring blood pressure,

Example:

- ... the attributable risk (...) is **greater** for coronary events **than** for stroke...
- This technique is **less** expensive and **more** convenient for patients **than** ABPM...
- **The more** blood your heart pumps **and the narrower** your arteries, **the higher** your blood pressure.

Conclusions

I have included my conclusions step by step on my *Quantitative and Qualitative analysis*.

Follow up

How will your syllabus be influenced by the outcome of your investigations?

I think my syllabus should focus on:

- **technical lexic** about high blood pressure, (such as *blood, pressure, hypertension, risk, heart, patient, measurement*, etc.), paying particular attention to **forms of nominalization** (such as *Measurement, Treatment, Monitoring*, etc.) and to **compound noun phrases** (such as *blood pressure measurement*, etc.)
- **specific language functions** (such as *Blood pressure measurement / Blood pressure can be /should be / is measured; To be at risk; Intervention(s) is/are based on...; Treatment is based on...*, etc. ;
- **verbal tenses**, particularly **Passive forms, Modal verbs** and **Imperative**;
- **comparatives**.

Simona Mosca