

Corpus Analysis

Tiziana Spanu

***The corpus analysed is composed of 7.474 words.**

QUANTITATIVE ANALYSIS

The corpus is based on material found on the Web, concerning **Medicine:**
” *HIV prevention*”.

The most frequent 100 words

Nr.	Word	Frequency
1.	and	291
2.	the	237
3.	to	229
4.	of	224
5.	for	179
6.	HIV	177
7.	be	116
8.	in	101
9.	or	100
10.	risk	97
11.	that	88
12.	a	84
13.	with	75
14.	transmission	70
15.	are	65
16.	infected	63
17.	patients	63
18.	can	59
19.	have	54
20.	health	54
21.	prevention	52
22.	should	49
23.	partners	48
24.	patient	48
25.	persons	48
26.	is	46
27.	screening	46
28.	their	44
29.	other	43

30.	by	41
31.	behaviours	40
32.	services	39
33.	not	38
34.	sex	36
35.	stds	34
36.	care	33
37.	who	33
38.	on	31
39.	these	28
40.	been	27
41.	behavioural	27
42.	clinicians	27
43.	interventions	27
44.	sexual	27
45.	drug	26
46.	might	26
47.	treatment	26
48.	as	25
49.	behaviour	25
50.	from	25
51.	information	25
52.	use	25
53.	has	24
54.	referral	24
55.	this	24
56.	counselling	22
57.	injection	22
58.	more	22
59.	at	21
60.	they	21
61.	about	20
62.	may	20
63.	medical	20
64.	partner	20
65.	related	20
66.	messages	19
67.	also	18
68.	however	18
69.	infection	18
70.	referrals	18
71.	will	18
72.	all	17

73.	recommendations	17
74.	risks	17
75.	s	17
76.	sharing	17
77.	studies	17
78.	viral	17
79.	among	16
80.	factors	16
81.	notification	16
82.	needle	15
83.	some	15
84.	such	15
85.	which	15
86.	clinician	14
87.	new	14
88.	reduce	14
89.	than	14
90.	abuse	13
91.	aids	13
92.	associated	13
93.	available	13
94.	department	13
95.	evidence	13
96.	including	13
97.	need	13
98.	provide	13
99.	testing	13
100.	address	12

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

Frequency list of adjectives

Qualifier	Frequency
infected	63
patient	48
other	43
behavioural	27
sexual	27
medical	20
related	20

viral	17
new	14
associated	13
available	13
including	13
Demonstrative	Frequency
that	88
these	28
this	24
Possessive	Frequency
their	44

Frequency list of adverbs

Adverbs	Frequency
by	41
not	38
as	25
also	18
however	18
all	17
some	15

Frequency list of conjunctions

Conjunctions	Frequency
and	291
for	179
or	100
as	25
than	14

Frequency list of nouns

Nouns	Frequency
HIV	177
risk	97
transmission	70
patients	63
health	54
prevention	52
partners	48
patient	48

persons	48
screening	46
behaviours	40
services	39
sex	36
stds	34
care	33
clinicians	27
interventions	27
drug	26
treatment	26
behaviour	25
information	25
has	24
referral	24
counselling	22
injection	22
partner	20
messages	19
infection	18
referrals	18
recommendations	17
risks	17
sharing	17
studies	17
factors	16
notification	16
needle	15
clinician	14
abuse	13
aids	13
department	13
evidence	13
need	13
testing	13
address	12

Frequency list of prepositions

Prepositions	Frequency
to	229
of	224
for	179
in	101
with	75
by	41
on	31
from	25

at	21
about	20

Frequency list of pronouns

Dimostrativi	Frequency
that	88
these	28
this	24
Personal	Frequency
they	21
Relative	Frequency
who	33
which	15
Indefinite	Frequency
some	15
all	17

Frequency list of verbs

be	116
risk	97
are	65
infected	63
can	59
have	54
should	49
is	46
care	33
been	27
might	26
use	25
has	24
may	20
related	20
sharing	17
reduce	14
abuse	13
associated	13
including	13
need	13
provide	13
testing	13
address	12

Frequency lists of articles

Definite	Frequency
the	237
Indefinite	Frequency
a	84

Core vocabulary:

Nouns	Adjectives	Verbs	Adverbs	Prepositions
hiv	that	be	by	to
risk	patient	risk	not	of
for	for	for	for	for
transmission	these	are	as	in
patients	behavioural	infected	also	with
health	sexual	can	however	by
prevention	this	have	all	on
partners	viral	should	some	from
patient	clinical	is	when	at
persons	social	care	though	about
screening	brief	been	often	among
behaviours	high	might	only	during
services	antiretroviral	use	after	into
sex	aware	has	although	against
stds	current	may	thus	
care	local	related	further	
clinicians	ongoing	sharing	less	
interventions	risky	reduce	any	
drug	safer	abuse	early	
treatment	often	associated	how	
behaviour	potential	including		
information	those	need		
has	effective	provide		
referral	important	testing		
counselling	national	address		
injection	active	change		
partner	additional	delivered		
messages	genital	increased		
infection	personal	antiretroviral		
referrals	sexually	done		
recommendations	specific	suggest		
risks	complete	transmitting		

sharing	difficult	used		
studies	general	concern		
factors	limited	concerns		
notification	noninfected	identifying		
needle	physical	include		
clinician	recent	preventing		
abuse		regarding		
aids		visit		
department		were		
evidence		becoming		
need		box		
testing		continue		
address		help		
reduction		informing		
social		known		
women		notified		
programs		obtained		
approaches		practices		
change		provided		
clinic		access		
exposure		addressed		
load		approach		
others		controlled		
serostatus		could		
state		discuss		
syringes		incorporating		
antiretroviral		made		
case		practice		
current		randomized		
departments		report		
idus		require		
local		asked		
person		culture		
recommendation		directed		
training		do		
community		ended		
concerns		identify		
condoms		increase		
office		limited		
problems		means		
questions		noninfected		
states		open		
table		prevent		

therapy		protect		
adolescents		setting		
age		was		
box				
disclosure				
effectiveness				
efficacy				
help				
history				
issues				
laws				
needles				
needs				
practices				
pregnancy				
resources				
settings				
staff				
access				
approach				
barriers				
educators				
Infections				
personal				
practice				
prophylaxis				
regulations				
report				
require				
strategies				
violence				
visits				
workers				
ability				
affect				
assessment				
benefit				
culture				
diseases				
drugs				
education				
equipment				
examination				

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

A nominalization consist of a transfer of the functions generally carried out by the verb to nouns and adjectives. Nominalization is often an economic alternative to a verb or a passive construction. The use of nominalizations enables complex information to be packaged into a phrase that is simple from a grammatical point of view and that can be picked up in the theme of the following sentence.

- Transmission
- Prevention
- Intervention
- Treatment
- Information
- Recommendation
- Sharing
- Notification
- Abuse
- Evidence
- Reduction
- Disclosure
- Regulation

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

• **HIV transmission**

(...Emphasis on prevention of transmission by HIV-infected persons...)

• (Infection can be reduced and transmission of HIV infection prevented...)

• (... A brief screening for HIV transmission risk behaviors..)

Prevention

• (...for incorporating HIV prevention into medical care of all...)

• (...patients for additional prevention interventions....)

• **Treatment**

• (...substantial advances in the treatment of human immunodeficiency...)

• (...not receiving antiretroviral treatment because they could be offered...)

- **Intervention**
- (...These disease intervention specialists can work closely...)
- (...is high without appropriate intervention, clinicians advised...)

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** (11) / **has been ...ed** (13)] and the **simple tenses** [**is/are ...ed** (46) or **was ...ed** (5)/**were ...ed** (8)]. They are used to explain the results or observations about studies or testing. They are used in particular when standard procedures are being described, and to describe the results of previous researches or procedures.

Examples:

We can find in the text the presence of the passive form:

11 Times have been, 13 times has been...

Es: (....informing partners that they have been exposed..)

(...They have been notified and tested for HIV...)

(...through heterosexual contact has been shown...)

(...a patient who has been treated with HAART for two years..)

8 times were...ed..5 times was..ed

(...partners were notified....)

(..The recommendations were developed by using an...)

(...This report was developed...)

46 times is...ed

(...is not infected...)

(..this report is not intended to duplicate..)

(..supporting the recommendation is indicated on a scale..)

(..for behavioral risk factors is not well defined...)

(...before the patient is seen by the clinicians...)

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

- **Passive forms**, are used to explain the results or observations on studies or testing about HIV diffusion.

They are used in particular when standard procedures are being described, and to describe the results of previous researches or procedures.

- **Modal verbs**

Can is used 59 times
Should is used 49 times
Might is used 26 times
Could is used 6 times
Would is used 2 times
May is used 20 times

They are used to give instructions or suggestions about HIV prevention. Modals, especially *may*, *might*, *could*, *would*, are one way of indicating the degree of certainty of a writer's commitment to a statement or a claim. So they are used to make cautious statements. This phenomenon is generally referred to as *hedging*.

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

We can find, for example,

- Prevention of **transmission** by HIV-infect....
- ..Can play an important role in HIV **prevention** messages
- ..A substantial advances in the **treatment** of human immunodeficiency...

9. **Some more information ...**

In the corpus analyzed we can find also an important presence of **comparatives** (**more** is found 22 times) and **logical connectors** (**however** (it is found 18 times), **moreover**, **therefore**).

They are important keys to understand the logical relationships in the texts, and therefore relevant to the teaching of reading, listening and writing in ESP.