

Corpus Analysis

Ugo Angelini

Domain : medicine

Topic : aids and adolescents

Total corpus words: 19.000

Quantitative analysis

What are the **most significant items** in that list?

The most significant items in this corpus were the nouns (79 out of 200), the adjective (39 out of 200) and the verbs (30 out of 200). Adverbs were not very common (5 out of 200)

Minimal core vocabulary

Most frequent nouns

[nouns] 79

1	hiv	264
2	aids	153
3	programs	136
4	prevention	128
5	program	126
6	school	72
7	students	68
8	use	65
9	pregnancy	62
10	health	60
11	education	53
12	teens	53
13	group	51
14	adolescents	50
15	people	47
16	teenagers	47
17	intervention	46
18	participants	44
19	one	43
20	active	40
21	community	39
22	information	37
23	sex	37
24	schools	35
25	age	34
26	behaviour	34

27	intercourse	32
28	study	32
29	behaviors	31
30	materials	31
31	year	31
32	cases	30
34	services	29
35	field	28
36	infection	27
37	skills	27
38	women	27
39	condoms	26
40	public	25
41	groups	23
42	partners	23
43	teenage	23
44	virus	22
45	number	21
46	report	21
47	drug	20
48	evaluation	20
49	knowledge	20
50	pasha	20
51	years	20
52	assistance	19
53	transmission	19

54	condom	18
55	family	18
56	know	18
57	men	18
58	sessions	18
59	ages	17
60	city	17
61	values	17
63	adolescent	16
64	black	16
65	counseling	16
66	peers	16
67	training	16
68	abstinence	15
69	care	15
70	communication	15
71	efforts	15
72	person	15
73	activity	14
74	assessment	14
75	comprehensive	14
76	development	14
77	peer	14
78	states	14
79	time	14

Most frequent verbs

[verbs] 30

1	are	124
2	is	95
4	have	76
5	was	75
6	be	72
7	were	67
8	had	47
9	reported	41
10	can	37
11	based	36
12	been	36

13	infected	30
14	has	29
15	including	25
16	conducted	24
17	designed	23
18	report	21
19	aged	20
20	should	20
20	know	18
21	tested	18
22	transmitted	17

23	compared	16
24	having	16
25	provide	16
26	said	16
27	used	16
28	may	16
29	found	15
30	include	14
31	using	14
32	would	14
33	must	13

Most frequent adjectives

[adjectives] 39

1	sexual	82
2	young	62
3	more	61
5	high	58
6	youth	55
7	one	43
8	active	40
9	two	32
10	infected	30
11	social	30
12	including	25
13	public	25

14	new	24
15	designed	23
16	many	22
17	some	22
18	each	21
20	aged	20
21	first	20
22	between	19
23	both	19
24	clinic	19
25	higher	19
26	national	19

27	behavioural	17
28	only	17
29	black	16
30	males	16
31	primary	15
32	activity	14
33	african	14
34	americans	14
36	comprehensive	14
39	medical	14

Most frequent adverbs

Adverbs [5]

1	Sexually	52
2	Also	39
3	Most	23
4	Only	17
5	Where	17

Qualitative analysis : Position and meaning of the most important items

NOUNS and ABBREVIATIONS

Word	Meaning	Concordance
PASHA	Program Archive on Sexuality, Health & Adolescence	#The Program Archive on Sexuality, Health & Adolescence (PASHA) is a new resource aimed at assisting practitioners around the country by assembling and disseminating materials from promising teenage pregnancy and STD/HIV/AIDS prevention programs
CDC	Centers for Disease Control and Prevention	#A recent Centers for Disease Control and Prevention (CDC) study found that about 40% of ninth graders,[...] report ever having coitus. # In recent CDC studies, adolescents who had coitus earlier in life reported a greater number of sexual partners # Since 1987, CDC has provided direct assistance to schools
HIV	Human immunodeficiency virus	# Now there are an estimated 42 million people living with HIV # HIV destroys a type of defence cell in the body called a CD4 helper lymphocyte #Half of all new HIV infections in the United States occur in people under 25 years of age
AIDS	Acquired Immune Deficiency Syndrome	# AIDS is one of the most serious, deadly diseases in human history # AIDS is caused by the HIV virus #the number of female adolescent AIDS cases has

		recently outnumbered male adolescent AIDS cases.
STD	sexually transmitted diseases	<ul style="list-style-type: none"> # Preventing HIV/STD Among adolescents # Among all sexually active people, teenagers have the highest rates of STD # Sexual intercourse places an adolescent at risk for HIV/STD
Prevention	To keep something from happening or existing	<ul style="list-style-type: none"> .# Easy access to HIV/STD health services is an important component of prevention. #Even though schools have many advantages in promoting HIV/STD prevention
Pregnancy	Having an unborn child in the body	<ul style="list-style-type: none"> #...funders and lawmakers are increasingly requiring those[interventions] aimed at preventing pregnancy, infection with sexually transmitted diseases #The Program Archive on Sexuality, Health & Adolescence (PASHA) is a new resource aimed at assisting practitioners around the country by assembling and disseminating materials from promising teenage pregnancy and STD/HIV/AIDS prevention
Infection	Passage of a disease from one body to another	<ul style="list-style-type: none"> # The human immunodeficiency virus known as HIV causes AIDS by infecting and damaging part of the body's defences against infection. # Babies born to mothers infected with HIV are also given special medicines to try to prevent HIV infection. # AIDS cases and HIV infection are increasing among adolescents
Condom	Rubber covering worn over male sex organ during sexual relations used as a mean to prevent pregnancy or as a protection against disease	<ul style="list-style-type: none"> # The proper and consistent use of latex condoms during coitus can greatly reduce a person's risk of acquiring or transmitting HIV/STD # The program aims to reduce

		infection by encouraging sexually active teenagers to use condoms . # At all activities, condoms are available, along with pamphlets explaining their correct use.
Drug	Synthetic or natural habit-forming substances which can provoke mental alterations in people who take such substances	#Sharing injecting drug needles can expose one directly to HIV # 61 high school students reported having injected an illegal drug .
Transmission	To send or pass from one person or thing to another	#Another activity targeting youth is CDC's Prevention Marketing Initiative (PMI), a large-scale social marketing program to influence behaviours that contribute to the sexual transmission of HIV and other diseases # Knowledge about HIV transmission was very good among teenagers in both schools
Virus	A living being even smaller than bacteria which causes infectious disease in the body	#AIDS is caused by the HIV virus (human immunodeficiency virus) #The only known way for the HIV virus to be transmitted from one person to another is when it is spread from the inside of an infected person's body to the inside of another person's body.

VERBS

WORD	MEANING	CONCORDANCE
Report	To give an account of something	<p># In recent CDC studies, adolescents who had coitus earlier in life reported a greater number of sexual partners. For example, 75% initiating coitus before age 18 reported having two or more partners,</p> <p># CDC reports that about 16% of 14-19 year olds reported four or more sexual partners.</p>
Infect	To transmit a disease	<p>#Thousands of U.S. teens become infected with HIV each year</p> <p># A person can become infected even if only tiny amounts of these fluids are spread.</p> <p>#The virus can be spread through an infected person's blood</p>
Transmit	To pass something(e.g. a disease) from a person to another	<p># The only known way for the HIV virus to be transmitted from one person to another is when it is spread from the inside of an infected person's body to the inside of another person's body.</p> <p># The sexually transmitted diseases, are an increasing health problem in our country.</p> <p># Research has shown that most Americans, including teenagers, understand how HIV is transmitted</p>
Conduct	To carry out, to perform	<p>#Research on effective behaviour change strategies and programs should be conducted.</p> <p># CDC studies conducted every 2 years in high schools indicate that approximately three-fourths of high school students have had sexual intercourse.</p> <p>#The study of the program was conducted with 87 black and Hispanic</p> <p>#The workshop was initially conducted with 209 undergraduate</p>
	The verbs I picked from my corpus- which are, I think, the most representatives- are generally used in the simple past tense, contrary to my expectations. Furthermore, they are frequently in the passive form, generally with the simple past or the infinite of the verb to be	

MODAL VERBS

Would	<ul style="list-style-type: none"> # Attention to improving overall health behavior, school performance, home and family life, and peer choice, for example, would be valuable. # Without PASHA, acquiring these materials and implementing the interventions would be difficult
Must	<ul style="list-style-type: none"> # HIV prevention programs must be culturally and specific population sensitive # Educational strategies dealing with prevention must be dealt... # Teachers must be given the support they need to tackle this dreadful information gap
Should	<ul style="list-style-type: none"> # Pregnant teens and women should be tested for HIV because women who receive treatment for HIV are much less likely to spread the virus to their babies. # Special attempts should be made to reach out-of-school youth # Research on effective behavior change strategies and programs should be conducted.
Can	<ul style="list-style-type: none"> # A person can become infected even if only tiny amounts of these fluids are spread. # HIV can be spread through sexual intercourse if one of the partners has the virus # Not all Youth can be reached through the school
May	<ul style="list-style-type: none"> # To reach teenagers and others not in school who may be at high risk for HIV infection # Class lessons may be reinforced through such activities as displaying posters and distributing wallet-sized HIV information cards # For users desiring more intensive, up-front training or on-site technical assistance, a section of the user's guide points to where and how these opportunities may be obtained.
	<p style="color: red;">Modal verbs are used here to give advice or to warn against something; again the passive form is very common with the modal verbs, generally with the infinite form of the verb to be</p>

To have

Has	<ul style="list-style-type: none"> # HIV can be spread through sexual intercourse if one of the partners has the virus. # The number of acquired immunodeficiency syndrome (AIDS) cases reported each year among U.S. adolescents (13-19 years Of age) has increased
Have	<ul style="list-style-type: none"> # 10% of teenage men tested for STDs have been found to have chlamydia.¹⁴ # Risk behaviour for HIV infection surveys have been generally made for specific groups # Two of the programs have been developed...

To be

Is	<ul style="list-style-type: none"> #AIDS is one of the most serious, deadly diseases #AIDS is caused by the HIV virus #A newborn baby is at risk of getting the HIV virus from his or her mother if she is infected #Since 1 in 5 reported AIDS cases is diagnosed in the 20-29 year age group
Are	<ul style="list-style-type: none"> # This can happen when infected fluid are passed from one person to another. A person can become infected even if only tiny amounts of these fluids are spread. # The sexually transmitted diseases, including human immunodeficiency virus infection and acquired immunodeficiency syndrome, are an increasing health problem # Some adolescents are at a higher HIV/STD risk # We know many more young people are infected with HIV
Was	<ul style="list-style-type: none"> #The median age of first sexual intercourse was about 16 years #The anonymity for both activities was guaranteed for all students #Odds Ratio (OR) was calculated...

ADJECTIVES

WORD	MEANING	CONCORDANCE
Sexual	Related to sex	<ul style="list-style-type: none"> # HIV can be spread through sexual intercourse # In recent CDC studies, adolescents who had coitus earlier in life reported a greater number of sexual partners. # The successful programs (1) had a narrow focus on reducing specific sexual risk-taking behaviors

Infected	Which/ who came in contact with an infection	# HIV- infected people # The only known way for the HIV virus to be transmitted from one person to another is when it is spread from the inside of an infected person's body to the inside of another person's body #In a single act of unprotected sex with an infected partner,
Behavioural	Connected with behavior	#Complemented with efforts to change behavioral norms #Representatives of affected populations, epidemiologists, behavioral scientists. #Positive behavioral impact

CONNECTORS

WORD	CONCORDANCE
Because (cause)	# Pregnant teens and women should be tested for HIV because women who receive treatment for HIV # Some adolescents are at a higher HIV/STD risk primarily because of societal conditions # We have made this option because the 8th year is the last year before they enter high-school.
However (contrast)	# Our criteria for effectiveness, however , err in the direction of overinclusion # However , on behavioural measures of sexual activity, there were no significant differences between the groups. # Very few, however , reported having injected drugs or shared needles.
As (cause)	# But as HIV destroys the lymphocytes, people with the virus begin to get serious infections # As other effective programs are identified and agree to participate, materials for these will be added to the collection.
Such as (example)	#This can happen when infected fluids - such as semen - are passed from one person to another #The number of adolescent cases of many STDs, such as chlamydia, #Also, using alcohol and other drugs such as crack cocaine impairs judgment
	The idea of contrast and explanation are important concepts for this topic, as it tries to explain in which cases HIV spreads, how and why, even though many conclusions scientists arrived at are liable to changes

- 1) Is there any **linguistic pattern** that is typical of the target language?
- 2) What are the most common forms of **nominalization** (if any)?
- 3) From the observation of certain items, can you detect the presence of specific **language functions** essential in the target language
- 4) Are there any **passive** forms? How often are they used and when?
- 5) What are the most **common verbal tenses** to be found?
- 6) Can you detect **compound noun phrases** that are typical of this variety, (i.e. nominal phrases)?
(Ex: The plant safety standards committee discussed recent air quality regulation announcements.)

I did not find either particular forms of nominalization or linguistic patterns, nor compound noun phrases typical of this area. On the contrary, I found many **abbreviations**, that, however are typical not only of ESP texts, but also of G.E texts, e.g.:

HIV Human immunodeficiency virus
 AIDS Acquired Immune Deficiency Syndrome
 PASHA Program Archive on Sexuality, Health & Adolescence
 CDC Centers for Disease Control and Prevention
 STD Sexually transmitted diseases

No **acronyms** in the whole corpus

As for the **language functions**, I focused my analysis on the connectors i.e., **because, as, however, such as** . These connectors indicate the necessity of the authors to give the idea of contrast, cause, advice and example. These concepts are necessary to explain in the whole text what AIDS and HIV are, how the virus spread and why.

- 7) Are there any passive forms? How often are they used and when?
- 8) What are the most common verbal tenses to be found?

The most common verbal tenses used in this corpus were the present simple and the passive. The past tense is generally expressed by using the present perfect. The verb list shows that the verb **to be** is the most frequently used in the corpus.

As for the modal verbs, **can** is the most frequently used, followed by **should, may, would** and **must**. Generally both **to be** and the **modal verbs** are used together with the past participle, then, the passive form is the most important verbal tense in this corpus. The use of the passive and of the modal verbs comes, I think, from the necessity to give advice to avoid contagion of HIV, to indicate how a prevention programme should be carried out, particularly in a school

Ugo Angelini