CREATION OF A CORPUS OF SPECIALISED LANGUAGE
TOPIC: BIOLOGY
ALESSANDRA LOI

QUANTITATIVE ANALYSIS

CORE VOCABULARY (AMONG FIRST 200 WORDS)

NOUNS		VERBS		ADVERBS		ADJECTIVES	
light	51	Is	142	As	63	this	59
reaction	49	are	107	however	15	photosynthetic	55
electron	47	be	57	even	14	other	38
time	45	have	43	about	13	which	37
photosynthesis	33	can	34	most	13	your	35
co2	32	has	30	but	11	these	29
energy	31	was	24	how	11	many	27
organisms	31	used	18	where	11	all	22
plants	28	been	17	out	10	such	21
transfer	27	use	17	there	10	their	19
center	26	process	16	when	10	some	17
oxygen	23	research	16	therefore	9	each	15
chlorophyll	20	do	15	also	8	several	14
pigments	20	wake	14	between	8	complex	12
protein	20	will	13	under	8	different	12
reactions	18	charge	11	after	7	genetic	12
bacteria	17	built	10	every	6	green	12
example	17	experiment	10			its	12
use	17	structure	10			high	11
membrane	16	found	9			higher	10
process	16	see	9			large	10
research	16	state	9			another	9
centers	15	absorbed	8			natural	9
temperature	15	called	8			organic	9
antenna	14	go	8			specific	8
wake	14	group	8			agricultural	7
day	13	make	8			any	7
proteins	13	reduced	8			genomic	7
systems	13	would	8			long	7
body	12	form	7			purple	7
complex	12	named	7				
entomology	12	occurs	7				
genetic	12	produce	7				
organism	12	resulting	7				
amount	11	determined	6				

charge	**11**						
clock	**11**						
dna	**11**						
fixation	**11**						
insects	**11**						
molecules	**11**						
water	**11**						
alarm	**10**						
algae	**10**						
experiment	**10**						
part	**10**						
plant	**10**						
production	**10**						
structure	**10**						
cells	**9**						
natural	**9**						
number	**9**						
organic	**9**						
pigment	**9**						
productivity	**9**						
separation	**9**						
state	**9**						
carbon	**8**						
carotenoids	**8**						
cell	**8**						
collection	**8**						
complexes	**8**						
enzyme	**8**						
group	**8**						
rates	**8**						
rhythms	**8**						
year	**8**						
chlorophylls	**7**						
concentration	**7**						
conditions	**7**						
cyanobacteria	**7**						
diversity	**7**						
figure	**7**						
form	**7**						
genomic	**7**						
period	**7**						
rate	**7**						
seconds	**7**						
subject	**7**						
yawns	**7**						
acid	**6**						
analysis	**6**						
bacteriochlorophyll	**6**						

MINIMAL CORE VOCABULARY (FIRST TEN WORDS)

NOUNS	VERBS	ADVERBS	ADJECTIVES
light	**Is**	**As**	**this**
reaction	**are**	**however**	**photosynthetic**
electron	**be**	**even**	**other**
time	**have**	**about**	**which**
photosynthesis	**can**	**most**	**your**
co2	**has**	**but**	**these**
energy	**was**	**how**	**many**
organisms	**used**	**where**	**all**
plants	**been**	**out**	**such**
transfer	**use**	**there**	**their**
center	**process**	**when**	**some**

QUALITATIVE ANALYSIS: OBSERVATIONS ABOUT THE LANGUAGE AND HOW IT IS USED

- **VERBS**

According to the verbs' frequency chart the 10 most used verbs are the following:
Is (142); are (107); be (57); have (43); can (34); has (30); was (24); used (18); been (17); use (17)

1) be

"Is" is the most frequent verb. It is used both as an ordinary verb and as an auxiliary followed by the past participle. The most frequent verb that goes with it is "used"; "are" is used both as an ordinary verb and as an auxiliary followed by a past participle. Only in some cases it is used in the present continuous form; "be" is preceded by modals and followed by past participles; "was" is used mainly in its passive form; "been" is basically used in the present perfect tense

2) have

"Have" is used both as an ordinary verb and as an auxiliary, but it is predominant the latter form, especially in the present perfect tense. "Has" is used as an ordinary verb but also as an auxiliary in the present perfect tense, both active and passive (12 passive forms out of 30 hits).

3) use of the imperative.

In the third article, where instructions are given to try out some very simple experiments, the imperative and simple present are prominent . Ex: *Get an oral…; Make sure you know…; Don't eat or drink anything….; Chart your body temperature with time Do you see a pattern? Ex: Is there a correlation? Is your reaction time faster or slower with warmer body temperatures?*

4) Modals

MODAL	HITS	MEANING/USE
Would	**8**	it is mostly used as a conditional and in hypothetical clauses (both 2nd and 3rd)
Will	**8**	it is mainly used as a future (to express that something will surely occur); only in a few cases in conditional clauses
Shall	**none**	
Should	**5**	Used differently: as a conditional giving advice and in hypothetical clauses
May	**8**	it is mainly used to express possibility and only in few cases risk and choice
Can	**34**	Used to express possibility; mainly used in the passive form: 22 hints out of 34
Could	**3**	It is differently used: as a conditional; to express possibility in the past and future in the past
Must	**4**	Used differently to express obligation and deduction in the past

- **NOUNS**

The nouns' frequency chart shows that the following are the 10 most used nouns in the articles (90 different nouns out of 200, not considering how many times they are repeated)

light (51) reaction (49) electron (47) time (45) photosynthesis (33) co2 (32) energy (31) organisms (31) plants (28) transfer (27)

In most cases these nouns are normally used as nouns. In other cases, they are used with an adjectival function and position. Examples:
1) normal use: ***carotenoids absorb blue-green light; physician who discovered that light is needed; which can donate an electron to the oxidized chlorophyll.***
2) adjectival function: ***Light energy; light intensities; light-induced charge separation; efficient electron transfer; the electron transport chain***
Some words are mostly used as nouns (time, experiment, process) whereas others are mainly used with an adjectival function (light and reaction. The word reaction is always used before center so to become: reaction center).

- **ADJECTIVES**

The adjectives' frequency chart shows that the following are the 10 most used adjectives in the articles (31 different adjectives out of 200, not considering how many times they are repeated):

this (59) photosynthetic (55) other (38) which (37) your (35) these (29) many (27) all (22) such (21)

What we can simply say about adjectives is that the demonstrative adjectives "this" is the most frequently used**.** We're dealing with explanations and descriptions, therefore the language focuses on showing, indicating, pointing at .

- **ADVERBS**

The adverbs' frequency chart shows that the following are the 10 most used in the articles (20 different adverbs out of 200, not considering how many times they are repeated). It should be underlined that not all the adverbs appearing in the list are always used with the functions of adverbs. Very often they are used as conjunctions or prepositions.

as (some out of 63) however (15) even (14) about (13) most (13) but (some out of) 11 (how 11) where (11) out (10) there (10)

1) the adverb "as" is used not very frequently as an adverb. Ex:… *collectively referred to as antenna*. On the contrary, it appears a lot as a conjunction or preposition.

2) "However" is used to comment on a previously stated fact with the meaning of "although sth is, was or may be true; nevertheless". Ex: *However, most pigments are not very…..;*

3) "Even" is mainly used with the meaning of " in spite of the fact or belief that; no matter whether". Ex: *Even though most plants are capable production process…..; …and even though they have improved.…..* * In this case, The Oxford Advanced Learners' Dictionary collocates the adverb "even" among the adverbs but with the function of conjunctions

4) "About" is always used with the meaning of "a little more or less than; a little before or after; approximately; nearly". Ex: … *generally has travelled about 2 nm to the other side of …..; has been estimated at about $117 billion per year*.

5) the adverbs "where" and "when" are normally used to introduce time and place clauses. Ex: *…take up CO2 during the night when the stomates are open; …….. synthetic organisms are found where the chloroplast has retained*

- **NOMINALIZATIONS**

The articles present a frequent use of nominalizations. They are present especially in the form of nominal phrases, that is, compound noun phrases. Example of nominalizations in our texts are:

1) first article: initial electron transfer charge (separation) reaction; water-splitting complex; light-induced charge separation; reduction-oxidation reactions; membrane-protein complex; light-induced charge separation. In this article the phrase "reaction center" appears 24 times

2) second article: research projects; terrestrial arthropod resources; collection-based systematic support

3) third article: reaction time; Graph paper; Stop watches; research subject; time estimation

- **FUNCTIONS**

As far as the language functions in our articles we have found:

1) use of **therefore** and **because** (not "as a result of") to underline cause and effect relationships
ex: *(…. of oxygen, because a close interaction with carotenoids is not always available under such circumstances. Therefore, all chlorophyll in a cell in aerobic organisms…..)*

2) use of expressions like "**such as**" to give examples, to illustrate. Scientific texts present examples. Ex: *…yet, other organisms, such as the purple and green bacteria,……*

3) use of **however**, **although**, **but, even though**. The discourse in our articles sometimes goes on with concessive conjunctions or clauses to express contrast, to express that some particular conditions may appear anyway. As if to state that often we cannot say that something is absolutely fixed or absolutely certain. There could always be exceptions. Conditions and results may be different.

- **HEDGING**

The following hedging devices appear in our corpus:

1) USE OF MODERATING ADJECTIVES such as "likely" (4 hits) and "apparent" (2 hits). Ex: *It is likely that the actual diversity of photosynthetic organisms……. it has become increasingly apparent*

2) USE OF MODERATING ADVERBS such as "apparently" (3 hits). Ex: … *which apparently are not easily targeted to and/or transported into the chloroplast….*

3) USE OF MODERATING VERBS such as "appear" (3 hits). Ex: *photosynthetic bacteria appear to have retained*

4) USE OF QUANTIFIERS to show generalizations or tendency such as "many" (some out of 27 hits), "few" (only one hit), "tend" (only one hit),"some" (17 hits). Ex: ……. *a single electron in many instances can be detected using……. has proven to have opened many doors in a variety of disciplines, ranging from biophysics to plant physiology……; very few organizations only one case; as they -for unknown reasons- tend to accumulate preferentially; Some experiments say that it does.*

Conclusions

The analysis of our Corpus revealed that nouns and verbs constitute the most used word class categories. Nouns are prominent in the sense that they appear in more different items than verbs; however, verbs are more common in the sense that they are more frequently repeated, especially as auxiliaries.

- NOUNS are normally used as nouns but also with an adjectival function and position

- As far as VERBS are concerned, "be" and "have" are predominant, since they are both used as ordinary verbs and as auxiliaries. The use of the passive form is massive and this is one of the most striking feature of scientific language. It is very frequent in the present perfect tense *(it has been proved, it has been found, they have been demonstrated, etc.*,). The present simple is normally the tense which expresses certainty and which is used to hold scientific statements which have been proved to be true. (ex. *Photosynthesis… is the process that converts energy in sunlight to chemical forms of energy …..).* In our first two articles we can state that both the present simple and the present perfect are the predominant tenses used to describe, to show, to illustrate the content of the articles. In the third article, the imperative and the simple present are the most common tenses, used to give instructions and to ask feedback on the experiments' results.
  Modals are rare in our articles. The only very frequent one is "can" used in its meaning of possibility.

- ADJECTIVES and ADVERBS are normally used and they do not present any particular use different from the everyday language. We're dealing with explanations and descriptions, therefore the language focuses on showing, indicating, pointing at.

As far as linguistic patterns that are typical of the target language, we can list the following:

- NOMINALIZATIONS

The articles present a frequent use of nominalizations. They are present especially in the form of nominal phrases, that is, compound noun phrases. They are used to nominalize ideas, to express ideas, concepts and actions with nouns. Nominalization is used when writing concisely and formally and, since our corpus of texts is a scientific one, this is the case.

- HEDGING **(**OR DISTANCING)

Hedging devices are used to take distance from what stated in a text and aim to convey the idea that there is no certainty. It is only suggested that something is possible, but there is no certainty. In our texts hedging is provided by moderating adjectives, moderating adverbs, moderating adjectives and quantifiers

- COMMON PHRASES
  photosynthetic organisms; Reaction Centers; electron transfer; infrared light; ultraviolet light; light energy; light reactions; photosynthesis process; photosynthetic systems; reaction time;

- SPECIFIC LANGUAGE FUNCTIONS They are examples of the rhetorical functions of : underlining cause and effect relationships (therefore and because**),** giving examples and illustrating (such as), stating that in spite of some conditions results may be different or change (however, although, but, even though**)**

Summing up, the main aims of our articles are: to describe, to illustrate, to explain. They illustrate researchers studies and results, as well. The third article gives instructions to try out some experiments about biological rhythms. The first two articles explain and illustrate what the photosynthesis process and entomology are. The use of the simple present and the present perfect are meant to explain something that actually has already been proven. But nothing is definite and science studies still go on leaving open many possibilities of further discoveries, as it is clearly stated in the first article: *This emphasizes that much relating to biodiversity and photosynthesis is still to be discovered, and that these discoveries are not limited to tropical rainforests and other ecological settings of large popular interest.*

Alessandra Loi