Corpus Analysis
Roberta Caddeo

Quantitative Analysis

NOUNS	ADJECTIVES	ADVERBS	VERBS
	These	When	Changes
	This	Over	May
	Much	Enough	Would
	Most	Mostly	Been
	Any	Just	Known
	Small	Yet	Increase
	several	after	Confirmed
	uv	probably	Make
	bright	then	Seen
	large	slightly	Account
	well	directly	Made
	global	before	Need
	last	ever	expected
	thin	inside	estimates
	another	potentially	had
	terrestrial	sometimes	increases
	martian	since	reach
	atmospheric	also	contain
	red	only	monitoring
	volcanic	so	record
	low	percent	seems
	planetary	on	exist
	unlike	how	increased
	climatic	about	surrounded
	industrial	even	documented
	least	too	making
	simple	still	provides
	third	out	have
	upper	ago	can
	able	away	will
	ancient	farther	warming
	beta	beneath	were
	colder	though	discovered
	distant	where	might
	greater	however	form
	new	never	observed
	active	once	know
	atomic	presently	maunder
	blue	through	produced
	changing	around	based
	good	now	mean
	higher	up	passes

	physical	almost	records
	radiative	likely	alter
	solar	perhaps	cooling
	other	below	do
	our	quite	find
	very	always	lies
	recent	within	remains
	less	down	should
	what	back	breathe
	same	often	cover
	early	again	enhanced
	their	moreover	feel
	larger		increasing
	longer		measured
	thick		occur
	great		added
	massive		began
	near		cannot
	each		contains
	Fahrenheit		include
	Little		pull
	those		is
	own		be
	dense		has
	hot		like
	infrared		was
	rocky		found
	above		see
	closer		called
	inner		could
	every		formed
	icy		affect
	internal		impacts
	poisonous		suggests
	sensitive		appears
	brighter		consists
	due		look
	hotter		orbiting
	late		went
	simply		does
	its		looks
	more		rise
	all		trapped
	long		did
	total		forcing
	two		including
	some		living
	different		received
	high		thought
	visible		believed
	solid		influences

	possible far half three few both largest many such magnetic first lower minimum smaller direct frozen closest dark overall average best chemical cold next significant entire gravitational huge polar right short current energetic heavy isotopic major maximum orbital outward surrounding powerful		projected brought caused indicate produce reaches reveal show think

MINIMAL CORE VOCABULARY

NOUNS	ADJECTIVES	VERBS	ADVERBS	CONNECTORS
System	Planetary	To monitor	When	Because
Atmosphere	Terrestrial	To confirm	Mostly	As
Climate	Martian	To estimate	Enough	While
Meteorite	Beta	To document	Yet	In addition
Greenhouse	Climatic	To record	Probably	Furthermore
Facula	Radiatiative	To find	Slightly	Then
Dioxide	Gravitational	To suggest	Potentially	However
Hydrogen	Fahrenheit	To cause	Always	Consequently
Radiation	Poisonous	To indicate	Perhaps	Moreover
Carbon	Isotopic	To reveal	Again	Certainly
Hemisphere	Infrared	To show	Likely	Though
Gravity	magnetic			
Particles				
Spacecraft				
Probe				
wavelengths				

QUALITATIVE ANALYSIS

In analysing the **vocabulary** of the most frequent words in my list, I divided it into the two ESP categories. In the first categories I stored some of the terms which has a high frequency in scientific and technical descriptions and discussions, specifying the frequency with which every single word occur in the text.

1): Planet (116), orbit (21), carbon (17), gases (15), dioxide (13), hydrogen (16), wavelength (13), gravity (10), climate (66), meteorite (4), satellite (4), pulsar (4), faculae (4), oxygen (13), greenhouse (26), methane (13), sunspots (8), terrestrial (11), beta (5), solar (174), Fahrenheit (8), infrared (7), gravitational (5) .

In the second category I stored the words whose meaning may vary across disciplines.

2): Cycle (22), energy (18), particles (7), systems (6), space, (21), probe (7), pressure (8), aerosols (6).

As regards the most frequent verbs occurring in the text, they are the ones commonly used to explain scientific phenomena or to formulate scientific hypothesis, such as: to confirm, to account,

to estimate, to monitor, to document, to record, to find , to suggest, to appear, to indicate, to reveal, to show.

I noticed a high frequency of the **modals** (can, may, should, might, could) which are used by the authors to report scientific hypothesis keeping the proper distance from them (hedging).

In my opinion, as the use of modals occur more frequently in the first of the three texts I selected, probably the author of the first one aimed at making his essay more appealing to the reader. In fact he often uses the modals to describe what would happen to the reader if he found himself on such or such other planet of the solar system.

On the other hand, the authors of the second and the third passage, making a quite limited use of the modals to express hypothesis, chose to report the demonstrated scientific facts, with the result of a more rigorous treatment of the argument.

I also noticed a wide use of the **comparatives**, which is related to the authors' will to describe the evolution of the phenomena diachronically through their development in time (ex.: "*The rings may have formed with Saturn itself, or they may have formed <u>much later</u>*"), and synchronically by comparing them to one another (ex.: "*Jupiter is <u>more massive</u> than all other planets and moons in our solar system*").

As regards the **collocations**, which describe the company that a word keeps, I only found some examples of concordances between noun and adjectives peculiar of the astronomic language, such as: *"<u>The rotational modulation</u> of solar total radiation...."*

"…<u>fairly rigid plates</u> on the planet's surface…."

"Uranus is bright enough to see with <u>unaided eye</u>"

There is a large number of examples of **nominalizations** along the text, in the form of transforming a verb form into a noun.

Ex.: *"The constant shifting of large, fairly rigid plates on the planet's surface…"*

"The burning of fossil fuels that has led to an increase in greenhouse gases…"

"It's a finding from nighttime astronomy that the present level of activity on the Sun is higher…"

"Detecting extra-solar planets directly is very difficult"

These are examples of how much is this feature favoured by academic writers, as it enables complex information to be packaged into a phrase that is simple from a grammatical point of view.

I also isolated a number of items which represent **specific language structures** typical of a coherent scientific discourse. For instance, there is a high frequency of cause and effect sentences, often marked by connectors such as "if" ("if" statements).

Ex.: *"If you went to Mercury…...…the Sun would look almost three times as large as it does from Earth…"*

"If you went to Uranus…..you could not breathe" because its atmosphere is poisonous, and you couldn't stand on the surface, because there isn't one"

Another connector I often met in the cause-effect sentences found in the text is "because".

Ex.: *"Because of this motion we can see the same side of Mercury each time the planet comes closest to Earth"*

"Because of Jupiter fast rotation rate, the clouds form bands of different colors, making Jupiter look like a striped beach ball with a big red spot in its southern hemisphere…"

This kind of sentence structure is typical of the scientific language of research, which is based on the principle of action-reaction and on the need of verifying every hypothesis that has been formulated.

I also noticed the presence of words as : furthermore, consequently, moreover, in addition, adding which have been useful for me to identify examples of the rhetorical function of sequencing the flow of a process or an argument and which contribute to the discourse coherence.

Ex.: *"Furthermore the overall long-term level of solar activity would have to fall…"*

"Consequently, because of these low temperatures and large gravity…"

"Moreover, what is known of the longer climatic record suggest that…"

"In addition, a steady stream of energetic particles…flow continuously outward from the Sun"

"Adding to the case is a finding from nighttime astronomy..."

I also have to mention the presence of **linguistic patterns** typical of the scientific or technical language, which generally tends to be more concise than any other. They are therefore used in order to express a long concept through a moderately short sentence.

Ex.: *"This long unanswered question has of late emerged anew, in the context of widespread concerns of impending greenhouse warming"*

"The century-long warming is a long anticipated sign of the climate system's response to human activities"

"…solar activity can drop from periods of many decades to levels that are lower than the minima of today's eleven-year cycle"

There is also a number of **compound-noun phrases** in my text, which aim at the same purpose of conciseness as the linguistic patterns mentioned above.

Ex.: *"Its atmosphere is streaked by high-altitude bright-white clouds"*

"The cloud-covered surface of Venus was long a mystery"

"The commercially-made compounds of chlorine, fluorine and carbon called halocarbons…"

As regards the verbs, **the verbal tenses** which most frequently occur in the text are:

-Simple present

-Present perfect

-Simple past

-Conditionals (often associated with the modals)

The passive form of the verbs can be often found throughout the whole text and deeply influences its structure. On the other hand the use of passive forms is typical of the scientific language, since authors aim at focusing the readers' attention on the cause of the phenomena they're explaining.

Ex.: *"The greenhouse effect on Venus is caused by its atmosphere"*

"The atmospheres of the terrestrial planets were created by three processes…"

"Uranus is encircled by very thin, dark rings"

CONCLUSIONS

As Register Analysis revealed, there is very little distinctive in the sentence grammar of my corpus texts beyond a tendency to favour some particular forms, such as the present simple tense, the passive voices, the conditional and nominal compound. They do not reveal any form that are not found in general English.

1- As regards the typical functions often found in my texts, I noticed the presence of a number of **nominalisations**, mostly in the form of transforming a verb into a noun in order to render a concept which it would have been difficult to express concisely in other ways.

I found some examples of **collocations** in the form of particular concordances noun-adjective.

I noticed the tendency throughout my text to the use some **linguistic patterns**, with the purpose of rendering the explanation of the scientific issues as concise as possible. That's why there is such a high frequency of elisions of conjunctions, of contractions of two or more words into one.

Other functions often to be found in my corpus aim at increasing the coherence of the text discourse, such as cause and effect sentences, marked by connectors as "if" or "because", and examples of the rhetorical function of sequencing the flow of an argument, marked by connectors as "furthermore", "consequently", "in addition", "adding to".

2- As regards the **nouns** most frequently occurring in my text, they are the ones specifically referring to the peculiar scientific field my corpus belong to, such as planets names, chemical elements, astronomic and atmospheric phenomena.

The same for the **adjectives**, which belong to the same semantic field and are used to qualify astronomic objects and phenomena.

As for the **verbs**, the most frequent ones are those referring to the formulation of scientific theories and hypothesis, such as: to confirm, to account, to estimate, to monitor, to document, to indicate and so on.

The **adverbs** are not so frequent throughout the text, and are mostly adverbs of time.

Ex.: "when", "just", "yet", "after", "before", "sometimes", "still", "never", "often", "always, "where".

C- The most common **verbal tenses** are: simple present (with the highest frequency), simple past, present perfect, conditionals. I think that the recurrence of the simple present is quite an obvious thing in texts which deal with scientific research that has always to be updated, while the use of the conditionals, often associated with the modals, indicates the authors' will to keep the distance from their hypothesis (hedging).

D- The most common **phrases** I found in my corpus are compound-noun phrases, used with the purpose to increase the conciseness of the scientific discourse.

FOLLOW UP

My work regarding the corpus analysis has been very useful for me, as it provided me with new instruments to deal with the language analysis. I think this will be very helpful for the

building of my syllabus as in my corpus analysis I had to focus my attention on the ESP language used in the texts I selected, and in particular on specific structures typical of the target language, such as nominalisations, collocations, the meaning of the use of modals, compound noun phrases, the use of the connectors and the way these structures have been used, which otherwise I would have neglected. Therefore now in building up my syllabus on ESP language I know on which features to focus the students' attention, which language structures are to be highlighted and practised in order to make them familiar with the specific technical language used in the scientific topic I chose.

Roberta Caddeo