CORPUS ANALYSIS

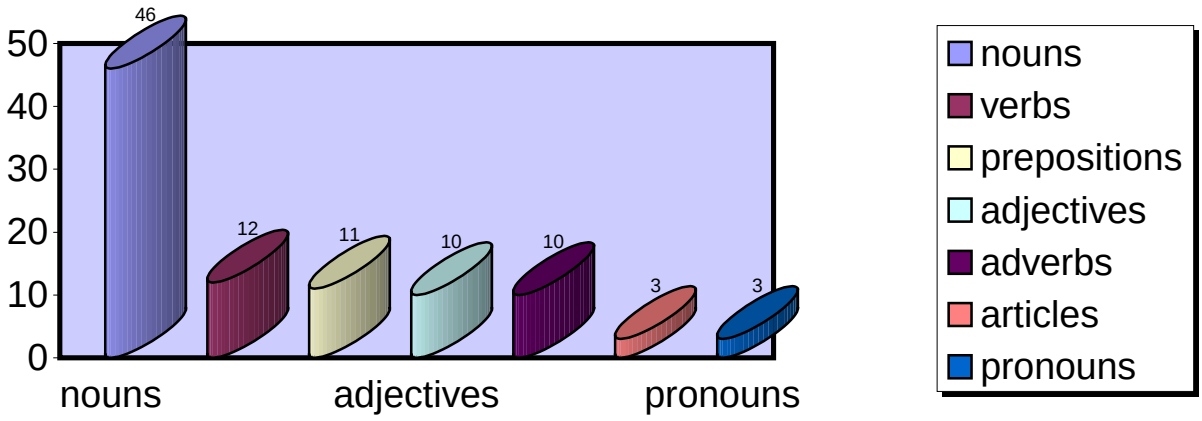
The language of Volcanology

Francesca Corrias

# Quantitative analysis

| | |
|---|---|
| Total number of **Word tokens** | **15234** |
| Total number of **Word Types** | **2846** |

**Distribution of the most frequent 100 words into grammatical categories**




The most significant items surely belong to the nouns category: volcano, lava, magma, eruption are some of the terms that are most related to the area of knowledge that is being considered. They will be analysed in detail during the qualitative analysis.

# Core Vocabulary

| NOUNS | | VERBS | | ADJECTIVES | | ADVERBS | |
|---|---|---|---|---|---|---|---|
| 1. clasts | 78 | 1. is | 70 | 1. dense | 29 | 1. as | 38 |
| 2. eruption | 73 | 2. are | 68 | 2. eruptive | 28 | 2. during | 3 |
| 3. magma | 60 | 3. was | 47 | 3. lower | 24 | 3. between | 27 |
| 4. crystal | 44 | 4. have | 33 | 4. vesicular | 24 | 4. only | 20 |
| 5. crystallization | 41 | 5. were | 30 | 5. large | 20 | 5. also | 17 |
| 6. groundmass | 36 | 6. show | 21 | 6. different | 19 | 6. generally | 14 |
| 7. density | 29 | 7. analysed | 15 | 7. glassy | 18 | 7. coarsely | 9 |
| 8. deposits | 29 | 8. like | 14 | 8. alkaline | 16 | 8. less | 9 |
| 9. feldspar | 29 | 9. present | 14 | 9. upper | 16 | 9. where | 9 |
| 10. eruptive | 28 | 10. suggest | 13 | 10. main | 15 | 10. highly | 8 |
| 11. sequence | 26 | 11. occurred | 12 | 11. high | 14 | 11. respectively | 8 |
| 12. table | 26 | 12. associated | 11 | 12. compositional | 13 | 12. very | 8 |
| 13. content | 24 | 13. characterized | 11 | 13. dark | 13 | 13. while | 8 |
| 14. size | 23 | 14. could | 11 | 14. first | 13 | 14. however | 7 |
| 15. ash | 22 | 15. increase | 11 | 15. grey | 13 | 15. mainly | 7 |
| 16. plug | 22 | 16. may | 10 | 16. similar | 13 | 16. when | 7 |
| 17. pumice | 43 | 17. using | 10 | 17. important | 12 | 17. clearly | 6 |
| 18. base | 21 | 18. bearing | 9 | 18. microvesicular | 11 | 18. often | 6 |
| 19. phase | 21 | 19. coloured | 9 | 19. dominant | 10 | 19. variably | 6 |
| 20. average | 20 | 20. due | 9 | 20. fragmentation | 10 | 20. always | 5 |
| 21. crystals | 19 | 21. formed | 9 | 21. irregular | 8 | 21. mostly | 5 |
| 22. dynamics | 19 | 22. represent | 9 | 22. its | 8 | 22. poorly | 5 |
| 23. fragments | 19 | 23. scattered | 9 | 23. low | 8 | 23. possibly | 5 |
| 24. vesicles | 19 | 24. supported | 9 | 24. rich | 8 | 24. probably | 5 |
| 25. activity | 18 | 25. been | 8 | 25. thick | 8 | 25. relatively | 5 |
| 26. textures | 18 | 26. melts | 8 | 26. larger | 7 | 26. significantly | 5 |
| 27. composition | 17 | 27. recorded | 8 | 27. linear | 7 | 27. irregularly | 4 |
| 28. features | 17 | 28. shaped | 8 | 28. stratigraphic | 7 | 28. largely | 4 |
| 29. can | 16 | 29. sized | 8 | 29. abundant | 6 | 29. progressively | 4 |
| 30. growth | 16 | 30. banded | 7 | 30. dry | 6 | 30. radially | 4 |
| 31. scoria | 16 | 31. cooling | 7 | 31. fine | 6 | 31. regularly | 4 |
| 32. blocks | 15 | 32. has | 7 | 32. grained | 6 | 32. slightly | 4 |
| 33. deposit | 15 | 33. occur | 7 | 33. intermediate | 6 | 33. moderately | 3 |
| 34. juvenile | 15 | 34. related | 7 | 34. minor | 6 | 34. nearly | 3 |
| 35. syn | 15 | 35. suggesting | 7 | 35. obsidian | 6 | 35. rather | 3 |
| 36. conditions | 14 | 36. crystallized | 6 | 36. progressive | 6 | 36. slowly | 3 |
| 37. data | 14 | 37. dominated | 6 | 37. shallow | 6 | 37. strongly | 3 |
| 38. days | 14 | 38. induced | 6 | 38. spherulitic | 6 | 38. commonly | 2 |
| 39. range | 14 | 39. melt | 6 | 39. volcanic | 6 | 39. early | 2 |
| 40. explosions | 13 | 40. represented | 6 | 40. white | 6 | 40. locally | 2 |
| 41. glass | 13 | 41. shows | 6 | 41. yellow | 6 | 41. occasionally | 2 |
| 42. population | 13 | 42. suggests | 6 | 42. ballistic | 5 | 42. plastically | 2 |
| 43. water | 13 | 43. close | 5 | 43. bent | 5 | 43. randomly | 2 |
| 44. coarse | 12 | 44. controlling | 5 | 44. contemporary | 5 | 44. almost | 1 |
| 45. conduit | 12 | 45. decrease | 5 | 45. continuous | 5 | 45. basically | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 46. feldspars | 12 | 46. increasing | 5 | 46. elongated | 5 | 46. completely | 1 |
| 47. material | 12 | 47. interpreted | 5 | 47. largest | 5 | 47. compositionally | 1 |
| 48. member | 12 | 48. measured | 5 | 48. major | 5 | 48. critically | 1 |
| 49. part | 12 | 49. performed | 5 | 49. rare | 5 | 49. differentially | 1 |
| 50. plagioclase | 12 | 50. selected | 5 | 50. single | 5 | 50. differently | 1 |
| 51. products | 12 | 51. suggested | 5 | 51. southern | 5 | 51. efficiently | 1 |
| 52. these | 12 | 52. according | 4 | 52. vulcanian | 5 | 52. eventually | 1 |
| 53. volume | 12 | 53. collected | 4 | 53. black | 4 | 53. frequently | 1 |
| 54. bands | 11 | 54. corresponding | 4 | 54. blocky | 4 | 54. gradually | 1 |
| 55. lapilli | 11 | 55. cover | 4 | 55. blundy | 4 | 55. incipiently | 1 |
| 56. mass | 11 | 56. described | 4 | 56. depositional | 4 | 56. laterally | 1 |
| 57. nucleation | 11 | 57. distributed | 4 | 57. fluid | 4 | 57. microscopically | 1 |
| 58. rate | 11 | 58. ejected | 4 | 58. general | 4 | 58. practically | 1 |
| 59. beds | 10 | 59. emplaced | 4 | 59. horizontal | 4 | 59. profoundly | 1 |
| 60. bombs | 10 | 60. evolved | 4 | 60. initial | 4 | 60. repeatedly | 1 |
| 61. dominant | 10 | 61. following | 4 | 61. long | 4 | 61. solely | 1 |
| 62. fragmentation | 10 | 62. forming | 4 | 62. magmatic | 4 | 62. steeply | 1 |
| 63. shapes | 10 | 63. preceding | 4 | 63. parallel | 4 | 63. totally | 1 |
| 64. surface | 10 | 64. ranging | 4 | 64. phonolitic | 4 | 64. univocally | 1 |
| 65. tuff | 10 | 65. should | 4 | 65. several | 4 | 65. vertically | 1 |
| 66. crater | 9 | 66. shown | 4 | 66. small | 4 | | |
| 67. eruptions | 9 | 67. sparks | 4 | 67. tabular | 4 | | |
| 68. flow | 9 | 68. considered | 3 | 68. thin | 4 | | |
| 69. gas | 9 | 69. considering | 3 | 69. volatile | 4 | | |
| 70. role | 9 | 70. consist | 3 | 70. analytical | 3 | | |
| 71. texture | 9 | 71. covered | 3 | 71. clear | 3 | | |
| 72. time | 9 | 72. cut | 3 | 72. common | 3 | | |
| 73. aggregates | 8 | 73. deformed | 3 | 73. consistent | 3 | | |
| 74. light | 8 | 74. driven | 3 | 74. double | 3 | | |
| 75. magmas | 8 | 75. estimated | 3 | 75. efficient | 3 | | |
| 76. occurrence | 8 | 76. existing | 3 | 76. elliptical | 3 | | |
| 77. properties | 8 | 77. explain | 3 | 77. euhedral | 3 | | |
| 78. viscosity | 8 | 78. found | 3 | 78. final | 3 | | |
| 79. distribution | 7 | 79. generating | 3 | 79. homogeneous | 3 | | |
| 80. hammer | 7 | 80. include | 3 | 80. intense | 3 | | |
| 81. layer | 7 | 81. indicates | 3 | 81. internal | 3 | | |
| 82. overpressure | 7 | 82. located | 3 | 82. negligible | 3 | | |
| 83. pressure | 7 | 83. make | 3 | 83. pargasitic | 3 | | |
| 84. process | 7 | 84. marked | 3 | 84. polygonal | 3 | | |
| 85. ranges | 7 | 85. mixed | 3 | 85. poor | 3 | | |
| 86. result | 7 | 86. mixture | 3 | 86. significant | 3 | | |
| 87. scoriae | 7 | 87. occurring | 3 | 87. skeletal | 3 | | |
| 88. shape | 7 | 88. owing | 3 | 88. slow | 3 | | |
| 89. solvus | 7 | 89. reported | 3 | 89. smaller | 3 | | |
| 90. system | 7 | 90. represents | 3 | 90. strong | 3 | | |
| 91. temperature | 7 | 91. resulting | 3 | 91. ternary | 3 | | |
| 92. thickness | 7 | 92. rising | 3 | 92. vertical | 3 | | |
| 93. trends | 7 | 93. scanning | 3 | 93. viscous | 3 | | |
| 94. vesicularity | 7 | | | 94. able | 2 | | |
| 95. behaviour | 6 | | | 95. absent | 2 | | |
| 96. crystallinity | 6 | | | 96. accretionary | 2 | | |
| 97. elements | 6 | | | | | | |

## A Minimal Core Vocabulary

| Nouns | Verbs | Adjectives | Adverbs | Prepositions |
|---|---|---|---|---|
| Eruption | Is | Dense | During | To |
| Volcano | Are | Eruptive | Generally | In |
| Magma | Was | Lower | Only | With |
| Clasts | Have | Upper | | From |
| Crystal | Were | Large | | By |
| Crystallization | Be | Vescicular | | At |
| Deposit | Has | Volcanic | | On |
| Density | Been | Different | | |
| Activity | Can | | | |
| Composition | Analysed | | | |
| Content | | | | |
| Cone | | | | |
| Lava | | | | |
| Dynamics | | | | |
| Fragments | | | | |
| Years | | | | |
| Flow | | | | |
| Degassing | | | | |

# Qualitative analysis

## 1. Some of the most frequent words: their use

**ERUPTION**

This word occurs 73 times. It is quite often accompanied by an adjective that qualifies the material, the effect on environment or on people, the time it occurs, or it is coupled with the name of a volcano:

Yellow Tuff eruption (material)
The AD 1538 Monte Nuovo eruption (volcano)
the July 1996 eruption (year)
cataclysmic eruption (effect)
It appears also associated to the word ‘dynamics’ (5 times).

**MAGMA**

This word occurs 60 times and it is interesting to observe its possible collocation:
It appears:
7 times with the word ascent (magma ascent)
6 times with the term chamber (magma chamber)
6 times with crystallization (magma crystallization)
5 times with the word plug (magma plug)
3 times with the word properties (magma properties)

**DENSE/DENSITY**: here is an interesting case of ambiguity. Indeed these words are used in two different ways:
In the majority of the occurrences Density represents a value that is the result of a mathematical operation, i.e. weight/volume. But, it can be sometimes used with a connotation that belongs to general English, and can mean, for instance, frequent, as in the example below:
*Crystal number density increases of two orders* where the word does not correspond to an exact numerical value like in *Average density is around 1.0x103 kg m-3*. At the same time the density value is often not specified, but the adjective and the noun refer most of the time to the operation mentioned above:

In *crystal-rich, dense clasts* the adjective, although it is not associated to a value, refers to the property of being **highly** dense.

**PUMICE**: this noun presents an interesting polisemy. Indeed it is used to describe a rock from two points of view: in a first classification a pumice is a clast (piece of rock) that is not smaller than 4 cm and is bigger than lapilli and ashes. But it is common to find another classification that is not dimensional but textural. Indeed a pumice is also very light and full of gas and so one can read of a pumice that is actually smaller than 4 cm. The example *ash-coated pumice lapilli* is quite clear, it would seem an oxymor because from a dimensional point of view a pumice cannot be a lapilli, but it can be so from a textural point of view.

## TO BE

The verb to be is highly frequent in the corpus that has been analysed. It is worth noting how the use of this verb varies according to the tenses in which it is expressed. Indeed, the present tense is used in the majority of occurrences to absolve the function of describing the physical properties. That is strictly concerned with the procedure that is used by researchers when they are reporting their investigations. There is always a section in which data are described physically and are defined from several points of view, i.e. dimension, weight, composition, age. Here are some examples:
*The juvenile fragments are dark coloured…*
*Vesicles are up to 1 cm…*
*Vesicles are irregularly shaped…*
*Groundmass crystals are mainly elongated…*
*The lithic component is always less than 2-3 wt%…*
*Mauna Loa is a "shield volcano"…*
*its summit is only 15km away from the town…*
The passive form is often used to report on what has been done by the researchers when they collected and analysed the data. This is typical not only of Volcanology research literature but of scientific texts in general. Here are some examples:
*bombs and blocks were repeatedly observed…*
*scoria-like clasts were studied…*
*Density measurements were performed…*
*Samples were dried at 60°C for 24 hours…*
*images were first made binary…*
*A suite of 100 clasts was randomly collected…*

## MODALS

There is a consistent use of modal verbs, in particular the verb can and the verb may. It is not volcanology specific but it is a common feature of every scientific text. It is not advisable to convey extreme certainty about one’s interpretation, so an

affirmation becomes less ‘heavy’ if it is expressed with modals: here are some examples:
*The LM sequence may be interpreted as…*
*The whole process can be envisaged as…*
*This dynamics could be related to the rapid…*
*signs that might indicate an increase…*

### 2. Linguistic patterns

**SUGGEST THAT**

There are 32 occurrences of this pattern and is used in the phase that commonly follows the description of data: the interpretation of data. It is used often impersonally but sometimes it occurs at the first person. The use of this pattern is relates to what has been said previously about the uncertainty or not peremptory aspect of scientists’affirmations :
*We suggest that the pre-eruption tempera…*
*We suggest however that initial water…*
*antirapakivi texture suggests that the pre-eruptive…*
*Our data suggest that the two explosions…*

### 3. nominalisations

*Degassing* (noun from verb), *to crystallise* (verb from noun), *crystallisation* (noun from verb that comes from noun), *glassy* (adjective from noun), *conversion* (noun from verb), *cooling* (noun from adjective), *zonation* (noun from noun).

### 4. nominal phrases

There is a very consistent amount of nominal phrases that can be also very long and complex. Here are some examples
*bread-crusted pumice bombs*
*pumice-bearing pyroclastic density currents*
*crystal number density*
*syn-eruptive degassing-induced crystallization*
*vapour-charged white clouds*

## Conclusions

Part of the conclusions have been already drawn in the qualitative analysis.
There are some points that have to be made clearer:

- in the corpus analysed there are two basic functions: the function of describing an item from a physical point of view and the function of interpreting the description. The text types that are used are therefore the descriptive and the argumentative.
- The most common verbal tense is the present simple**.** followed by the simple past. But there is a great amount of passive forms as well. This, as it was mentioned before, is a common characteristic of scientific English, but not only. It can be noticed that even in this analysis I exploited the passive form to convey a certain neutrality to the text.
- Apart from the auxiliaries to be and to have, which are the most frequent verbs in the texts analysed, there is a great amount of verbs that pertain to the semantic area of the subject studied, i.e. volcanology: To deform, to erupt, to crystallize. Some of these verbs, though, are quite common in all scientific varieties of English. This is the case of to measure, to increase/decrease, to eject, to deform.
- There were not a very large quantity of linguistic patterns or formulas in the corpus. One that was found was to suggest that, but one can still recognize a patterned use of some particular verbs (like to consider or to show).
- There are a great amount of very complex noun phrases, as it was mentioned above, and this is quite typical of every scientific text, but it is also a characteristic of the English language itself. The use of the noun phrases is also a way to be concise and to concentrate a great number of information.

Francesca Corrias