

Key words and key phrases in a corpus of travel writing

From Early modern English literature to contemporary "blooks"

Andrea Grebig

presentation by Libor Vaněk

November 23, 2011

Content

- 1 Introduction
- 2 Key words: Implications and other studies
- 3 The corpus
- 4 Method and Findings

Goal

Goal

Show the changing roles and implications of travelling in society over a considerable time span.

How?

- key words
- key-key words
- associates
- multi-word units
- most frequent 5-grams

Content

- 1 Introduction
- 2 Key words: Implications and other studies**
- 3 The corpus
- 4 Method and Findings

First studies

J. R. Firth

The technique of semantics, *Transactions of the Philological Society* (1935)

- Suggested investigating the “distribution of sociologically important words, what one might call focal or pivotal words in all their derivatives and compounds in sociologically significant contexts”

Finding Key words

R. Williams

Keywords, 2nd edn. (1983)

- keywords =
 - “significant, binding words in certain activities and their interpretation”
 - “significant, indicative words in certain forms of thought”.
- stresses **individual choice**.

M. Scott

PC analysis of key words – and key key words, *System* (1997)

- “**statistical** method of selecting words according to their relative frequency”
- independent of personal views and preferences

Content

- 1 Introduction
- 2 Key words: Implications and other studies
- 3 The corpus**
- 4 Method and Findings

Corpus

- Travel writing from 16th to 21st century
- Each century covers about 500.000 words
- 21st century: texts published on the internet
 - well-structured and well-edited platform
 - reports of a coherent format, more like short-stories
 - platform maintained by a chain of shops providing the travellers with all necessary equipment
 - no marketing / advertising objectives

Content

- 1 Introduction
- 2 Key words: Implications and other studies
- 3 The corpus
- 4 Method and Findings**

Key words: method

- relative frequency of words
- comparison with background corpus
 - Scott & Tribble (2006) suggest a background corpus for comparison at least five times the size of the investigated text
 - C21 to C19: a self-compiled three-million word background corpus of mixed written and spoken sources
 - C18 to C16: section of the Helsinki Corpus, 551,000 words
- 100 most frequent key words for each century, further intuitively grouped into sets of semantically related words

Key words: results, example I.

In the more recent three centuries investigated, **references to countries** are more **varied** and more **frequent**.

- C16: Russia, England, English, Persia, Moscow, Tartars, Spaniards
- C17: Persia/-n/-s
- ...
- C20: Taiwan/ese, Ital/y/ian, Calabria/n, Patagonia, Naples, Rio, Mandarin, Chinese, Buenos Aires, Ionian, Albanian, Chilean, Hellenic
- C21: Sydney, Australia, Bangkok, Thai/land, Korea/n Belize,

Key words: results, example II.

Decline and **disappearance of boat travel.**

- C16: island/s, river, ship/s, sea, land, wind, captain, voyage, mariners, sailed, coast, journey, north, east, west, northwest, southwest, south, southeast, northeast
- C18: island/s, board, latitude, longitude, sea, shore, captain, sail, bay, cape, boat
- C19 - C21: almost no key word considering sea travel

Key words: results, example III.

Change of background corpus can bring out characteristic between the text under investigation and some “norm” where, obviously, today’s norms are different from 18th century norms.

Comparison **C18** with **Helsinki Corpus**:

- island/s, board, latitude, longitude, sea, shore, captain, sail, bay, cape, boat

Comparison **C18** with background corpus of **current English**:

- latitude, longitude, sea, island, shore, W(est), E(ast), captain, boat, bay, N(orth), voyage, cape, board, isle/s, canoes

Key words: results, example IV.

More **spoken style** in current century.

- C21, pronouns: we, my, I our, us, me (top of list)

Key-key words, Associates: definitions

Key-key words

- words that are found to be key words in a large number of individual texts
- lexical choices which are typical for the genre

Associates

- key words that are found to co-occur significantly with a key-key word
- profile of the contexts of words and by extension also of the texts

Key-key words: example I.

The most frequent key-key words in C21:

- guy, guys, locals, tourist, tourists, bag, bags, backpackers, backpack, travelers
- tour, trip, road, walk, walking, ride, driver
- us, we, my, I, me, our, everyone
- water, beach, hotel, village, town, city, center, shops, stay, stayed, night
- up, around, back, off, to, along, after, out, headed, few, day, a next
- breakfast, food, chicken, banana, lunch
- bus, taxi, boat, guide, hike
- hour, hours, minutes, awhile
- huge, didn't, stopped, decided, toilet

Key-key words: example II.

The most frequent key-key words in C20:

- hills, hill, place, spot, stones, sea, mountain, mountains, valley, land, landscape, trees, earth, sun, sunshine, earthquake
- green, red, black, little, remote, dim, charm
- city, village, villages, town
- me, my, myself, its, they, their, them, his, some, every
- distant, among, amid, hither
- beast, creatures, folks
- old, a, once, discovered, like, with, into, face, dead, and, ride, was, fashion, yet, had

Key-key words: conclusion

Comparison C21 with C20:

- C21: travellers or rather, backpackers, concern with the daily needs of moving around, staying overnight, eating and meeting people
- C20: more concern with details of the things seen and places travelled to.
- C20: no informal use such as guy/s.
- C20: no reference to fellow travellers or tourists
- C20: coverage seems to be more varied (higher number of less frequently occurring key-key words than in C21)

Associates: example

Tourists, travellers or backpackers:

- ride, road, trip, driver, beach, tour, walking; guy, guys; off, to, next, up; my; day

Tourists:

- beach; around; us

Travellers:

- walk; me; out; water

Backpackers:

- guide, town; night, hours; everyone; food

Extended lexical units: description

- Author picked one intuitively interesting key-key word from C17 - **abundance** - and looked at its contextual realisation through the centuries
- *abundance* = "when there is more than enough of something" (Cambridge Dictionary)

Extended lexical units: results, example.

C17: key-key word **abundance**:

- 87 occurrences
- 70: abundance of [optional classification / adjective] (concrete) noun
- 14: concrete noun [verb phrase] in [great/er] abundance
- 3: abundance (without preceding proposition or following of)

Extended lexical units: results.

- **decline of the use** of the node abundance from the earlier to the later centuries
- development in distribution from a preference of the structure **abundance of to in abundance**
- overall **more frequent in travel literature** than in other text types (inherently descriptive aspect)

Key phrases: description

- most frequent **5-word phrase-frames** over the centuries
- *phrase-frame* looks for recurring structures that vary in one slot.

Key phrases: example I.

C16:

- **in the * of the**: 65 occurrences, 14 variants
* = (midst 11, name 9, middle 7, time 6, absence 5, spring 4, bottom 4, presence 4, place 3, middes 3, dominions 3, midst 3, mouth 3)
- **at the * of the**: 33 occurrences, 8 variants
* = (mouth 7, discretion 5, time 5, end 4, sign 3, beginning 3, charge 3, request 3)

Key phrases: example II.

C21:

- **in the middle of ***: 64 occurrences, 4 variants
* = (the 38, nowhere 13, a 10, frikkin 3)
- **in the * of the**: 60 occurrences, 6 variants
* = (middle 38, back 6, shade 5, center 4, centre 4, heart 3)
- **at the * of the**: 48 occurrences, 7 variants
* = (end 16, front 7, top 7, foot 6, side 5, edge 4, base 3)

Key phrases: example III.

The most frequent phrases in C21 and C20

- preposition-article-noun-of-article
- the noun slot is mostly filled with fairly general spatial terms.

The most frequent phrases in C19 and C18

- preposition-article-noun-of-article
- realization of the noun slot changes:
 - the spatial references are more concrete
 - temporal references increase
 - some metaphorical uses

Any questions?

Thank you!