Text mining of TripAdvisor visitor attraction reviews –
The cases of Hammershus on Bornholm island and Visby City Wall

By Carl Henrik Marcussen, senior researcher, PhD

Centre for Regional and Tourism Research, www.crt.dk,

@ "Sustainable Tourism in a Digital World" conference,

24th-26th Sept. 2019, University of Uppsala Gotland campus

Overview

1. Introduction – research question

2. Related literature

3. Methodology, technique

4. Results: Hammershus castle ruin + visitor centre - Visby City Wall - and comparison

5. Conclusion / Q & A / Discussion

1. Introduction – cf. the conference theme

Sustainable Tourism …. in a Digital World

Hammershus

Visby City Wall

Tripadvisor
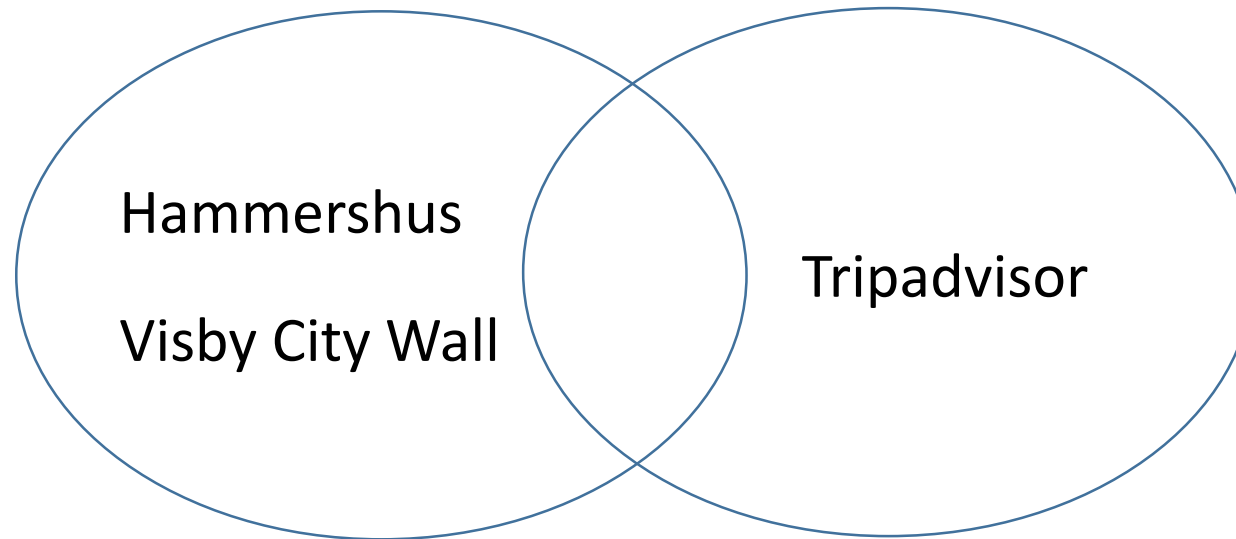
1. Introduction – research question

Background: Why text mining? Why the two sites?

Research question(s):
What are the association between words and review score and among the words use for describing selected visitor attractions on two Scandinavian island? **Which words are used significantly more** frequently about either of the two historic sites, Visby City Wall and Hammershus castle ruin including the associated new visitor centre?

2. Bornholm and Gotland
.. Islands in the Baltic Sea

Twining-Ward and Baum (1998):

"The impressive castle of **Hammershus** on the island of **Bornholm** belonged to the Archbishop of Lund".

"The top cultural attractions of the island include the traditional fish smoke houses, the dramatic **Hammershus castle**, Bornholm art gallery as well as the pottery and glassblowing workshops.

Gotland's main attraction is the **Hanseatic walled city of Visby**, which has been greatly restored and recently designated a **UNESCO** World Heritage Site."

2. Text mining …. and related concepts

- Text mining /
- Text data mining
- Text analytics
- Text analysis
- Sentiment analysis

2. Text mining …. example of procedure

1. Import event data into mining tool.
2. Linguistic processing.
3. Factor analysis
4. Cluster analysis.

Reference: Kaushik and Naithani (2016)
A Comprehensive Study of Text Mining Approach[es].

3. Text mining – methodology, technique

A. Acquiring the data:

Phython, Google Chrome, cut-and-paste

B. Analysing the data:

Special software ... or standard tools

3. Methodology and data - Procedure

- Collect the reviews – or other text/data
- Translate into English – if necessary
- Combine all reviews into a single dataset
- Conduct frequency count of (all) words
- Converting text into numbers (0/1)
- Analyse the reviews (for each attraction combined and separately)

4. An initial results: Seasonality
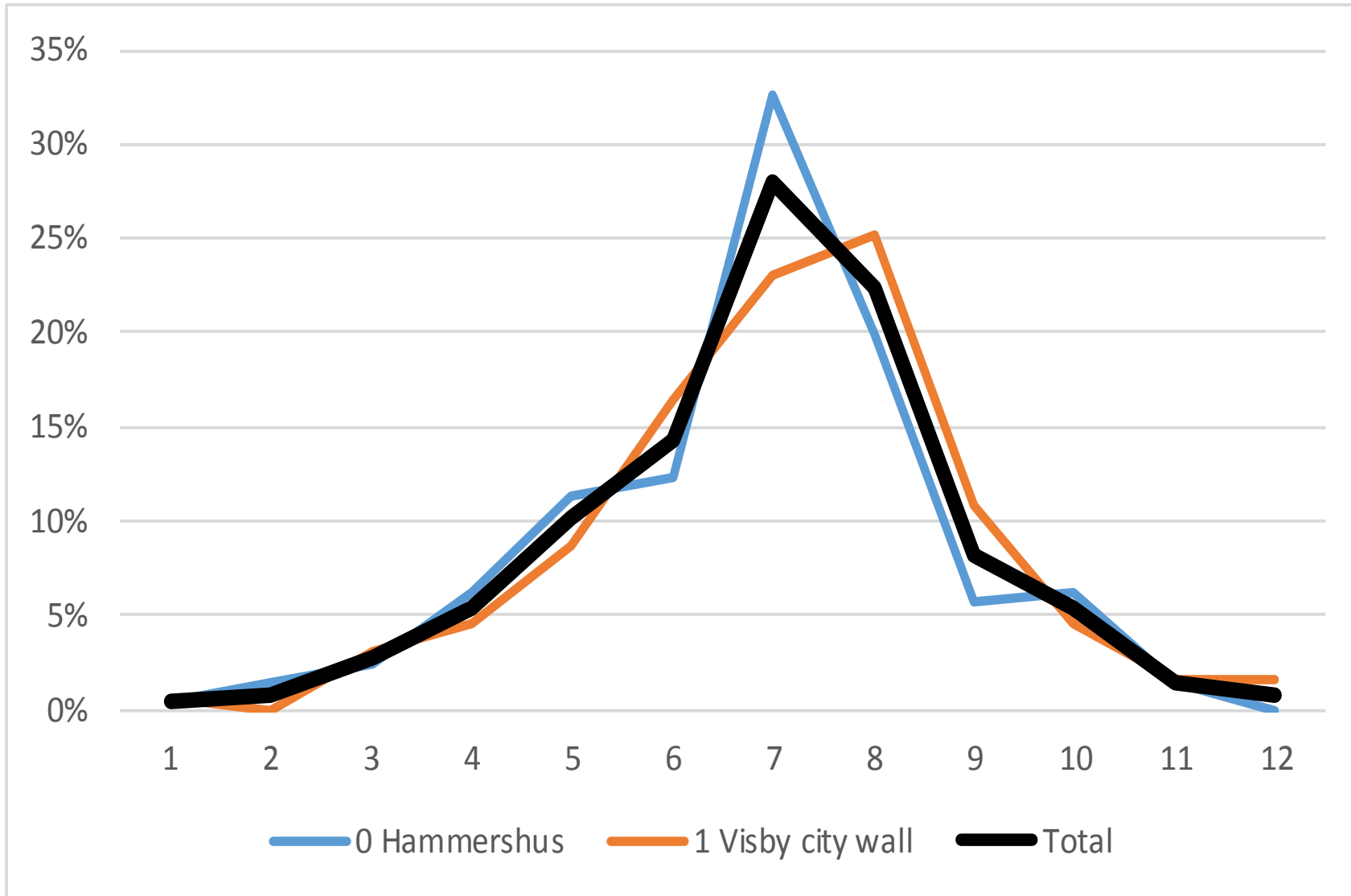
4. Positive or negative effect on score: Hammershus castle ruin and visitor center

Model	Determinants	B	t	Sig.	VIF
1	(Constant)	**4,569**	40,578	0,000	
	from	**0,278**	4,108	0,000	1,085
	and_	**0,293**	3,939	0,000	1,143
	hours	**0,439**	3,381	0,001	1,184
	it	**0,192**	3,121	0,002	1,111
	experience	**0,304**	2,846	0,005	1,086
	visited	**0,348**	2,733	0,007	1,221
	explore	**0,587**	2,630	0,009	1,126
	inside	**0,656**	2,621	0,010	1,068
	ruins	**0,187**	2,559	0,011	1,133
	trip	**0,269**	2,389	0,018	1,077
	enjoyed	**0,491**	2,372	0,019	1,443
	for	**0,170**	2,326	0,021	1,227
	week	**1,171**	2,261	0,025	1,540

Model	Determinants	B	t	Sig.	VIF
1	(Constant)	**4,569**	40,578	0,000	
	located	**-0,287**	-2,064	0,040	1,063
	view	**-0,153**	-2,150	0,033	1,162
	being	**-0,338**	-2,230	0,027	1,138
	over	**-0,299**	-2,383	0,018	1,187
	one	**-0,256**	-2,408	0,017	1,073
	the	**-0,295**	-3,050	0,003	1,143
	had	**-0,405**	-3,382	0,001	1,152
	into	**-0,505**	-3,487	0,001	1,153
	however	**-0,596**	-4,315	0,000	1,146
	fortress	**-0,651**	-4,401	0,000	1,087
	but	**-0,335**	-4,530	0,000	1,113
	long	**-1,447**	-5,105	0,000	1,819

Note: Dependent Variable: Score.   R2 adj.=0.515

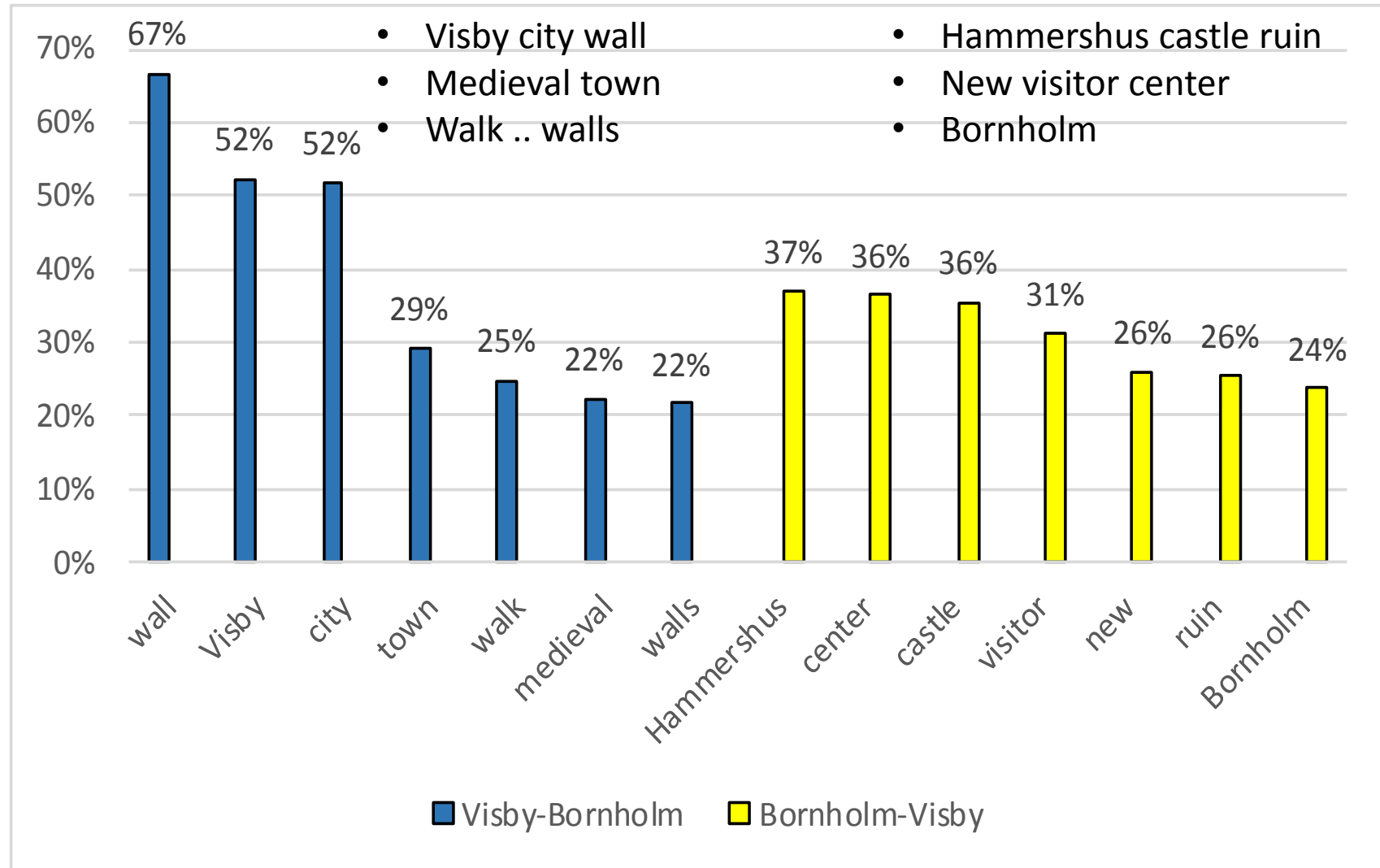Selecting only cases for which Hammershus_and_VC = 1.

4. Positive or negative effect on score: Visby City Wall – UNESCO sight

Most scores are 5 out of 5. – Average: 4.75.

So, it's hard to differentiate between words affecting the review score positively or negatively (low R2 adj., 0.087)

"Inside" is one of the plus words for Visby City Wall.

Main differences in proportions of word occurrence between Visby city walls and Hammershus in TripAdvisor reviews

- Visby city wall
- Medieval town
- Walk .. walls

- Hammershus castle ruin
- New visitor center
- Bornholm

Chart data:
- wall: 67%
- Visby: 52%
- city: 52%
- town: 29%
- walk: 25%
- medieval: 22%
- walls: 22%
- Hammershus: 37%
- center: 36%
- castle: 36%
- visitor: 31%
- new: 26%
- ruin: 26%
- Bornholm: 24%

Legend: ■ Visby-Bornholm ■ Bornholm-Visby

Selected visitor attractions at Tripadvisor

Attraction	Reviews	Score	Country	Province
Visby City Wall	551	4,75	Sweden	Gotland
Gotland's Museum	439	4,52	Sweden	Gotland
Kalmar Castle	771	4,48	Sweden	Blekinge
Borgholm Castle	244	4,16	Sweden	Blekinge
Glimmingehus	130	4,11	Sweden	Scania
Hammershus	520	4,52	Denmark	Bornholm
Østerlars Round Church	297	4,33	Denmark	Bornholm
Kronborg Castle	2887	4,45	Denmark	North Sealand

Selected visitor attractions at Tripadvisor
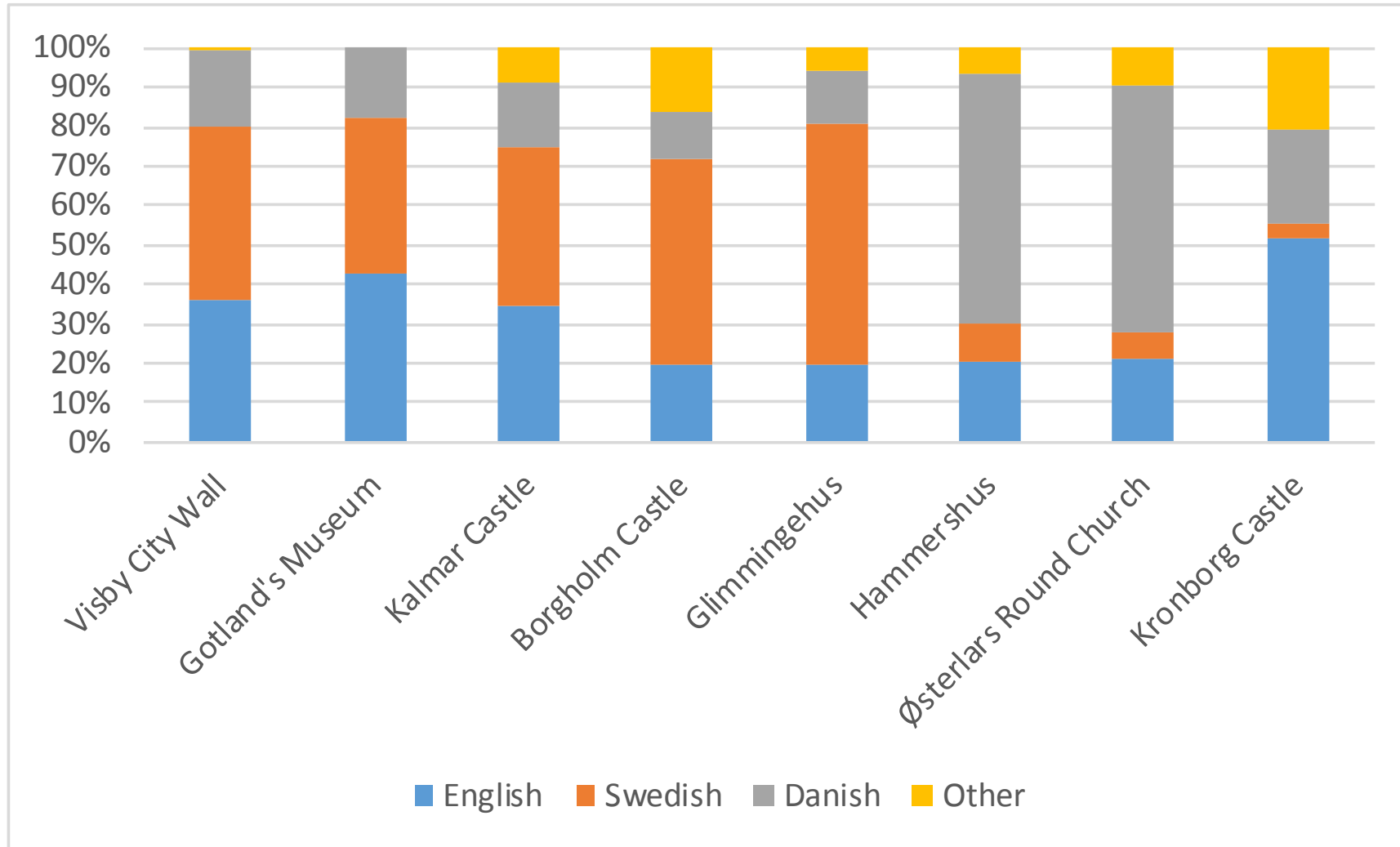
Words used significantly more often about **Hammershus** than about Visby City Wall

1-10	11-20	21-30	31-40	41-50
01. castle	11. views	21. beautiful	31. if	41. area
02. center	12. at	22. from	32. visited	42. located
03. visitor	13. nature	23. here	33. northern	43. super
04. new	14. top	24. location	34. sunny	44. were
05. ruin	15. parking	25. building	35. being	45. not
06. Bornholm	16. however	26. great	36. now	46. sea
07. ruins	17. Danish	27. but	37. coast	47. exciting
08. view	18. family	28. exhibition	38. little	48. trip
09. place	19. surroundings	29. must	39. with	49. for
10. free	20. good	30. scenery	40. windy	50. hours

Words used significantly more often about **Visby City Wall** than about Hammershus

1-8	9-16	17-24	25-32	33-38
01. wall	09. outside	17. streets	25. climb	33. way
02. Visby	10. old	18. heritage	26. of	34. week
03. city	11. Gotland	19. well	27. how	35. many
04. town	12. towers	20. walled	28. much	36. makes
05. walls	13. along	21. UNESCO	29. part	37. whole
06. medieval	14. around	22. cruise	30. walked	38. points
07. preserved	15. inside	23. ancient	31. days	
08. walk	16. within	24. intact	32. during	

→ Aspects of sustainability

195 words used at least five times about Hammershus – divided into 72 (78) groups by means of factor analysis – 11 first shown

1) new center visitor into 2) it_is its
3) thirteenth century 4) spent over two2 hours
5) we were visited 6) guided tour spend
7) must visiting when Bornholm 8) has been since
9) I could 10) Baltic Sea 11) definitely worth visit
years ……..

Dendogram for words typically used about Visby City Wall (top third) and Hammershus (lower part) in TripAdvisor reviews
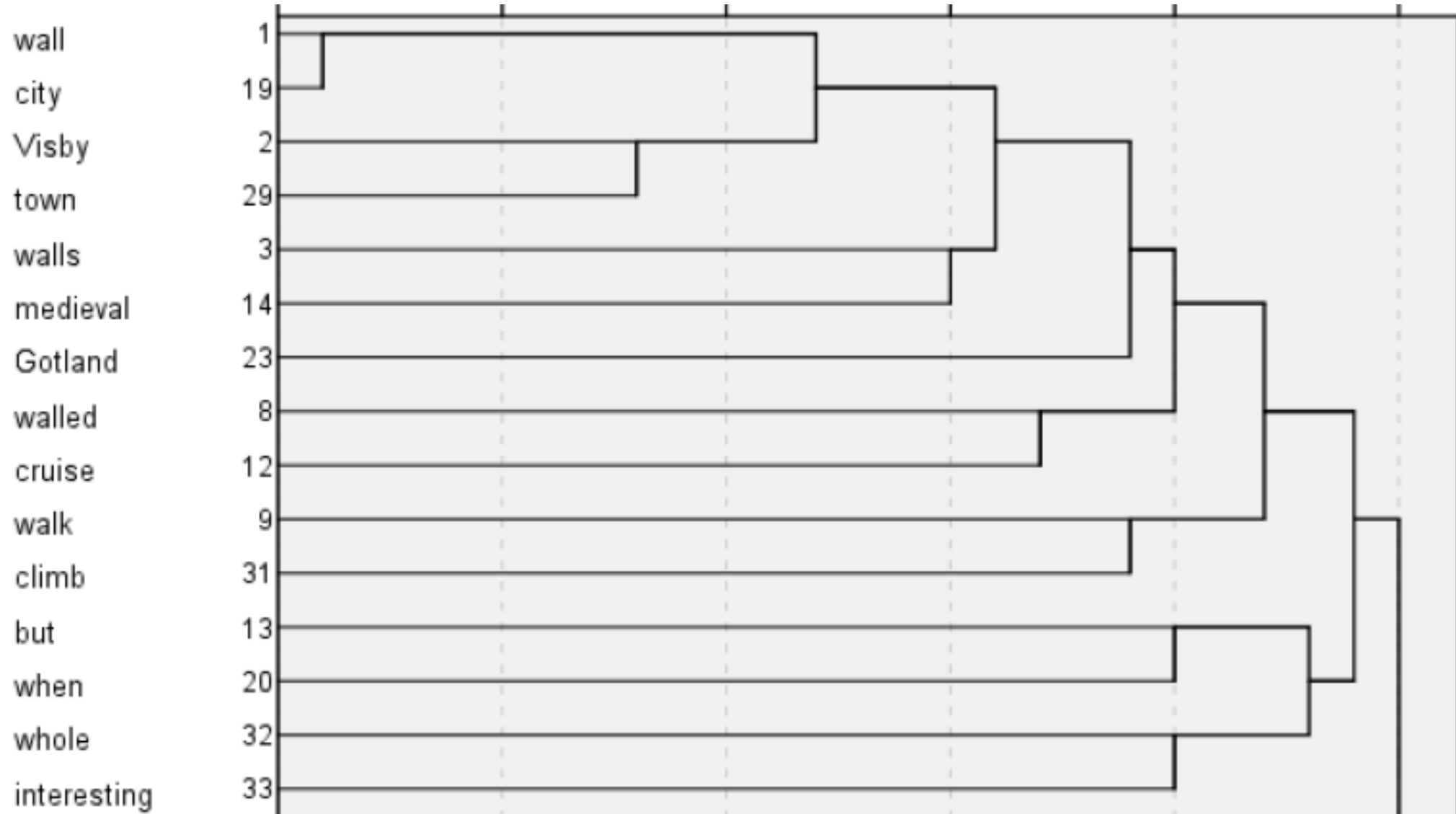
38 words used significantly more frequent about Visby City Wall than about Hammerhus – divided into 17 groups by means of factor analysis

Compo-nent	Word	Mean	Corre-lation	% of variance expained	Label
1	preserved	68%	0,898	5,00	well preserved
	well	52%	0,894		
2	UNESCO	52%	0,877	4,61	UNESCO world heritage
	heritage	30%	0,840		
3	week	23%	0,814	4,53	during medieval week
	during	27%	0,746		
	medieval	19%	0,411		
4	inside	41%	0,820	4,48	inside and outside
	outside	13%	0,809		
5	climb	27%	0,825	4,42	climb towers
	towers	9%	0,676		
6	walls	10%	0,750	4,17	wall or walls
6x	wall	13%	-0,737		
7	cruise	33%	0,684	4,07	cruise
	part	11%	0,537		
7c	days	7%	-0,352		
8	around	7%	0,794	3,98	walk around
	walk	6%	0,599		
9	old	28%	0,698	3,91	within old town
	within	5%	0,619		
	town	5%	0,585		

Conclusion

- Standard analytical tools are applicable to analysis of words and review scores, or between words – when words are converted to numbers.

- Among descriptive analytical tool such as factor analysis, cluster analysis and multidimensional scaling, in particular factor analysis (PCA) is a powerful tool.

- At least for small scale scraping of data, Google Chrome is applicable.

- Several tools can be used to test statistical significance, in text mining – but they have their limitations, advantages and disadvantages

Wordcloud, Hammershus, Tripadvisor
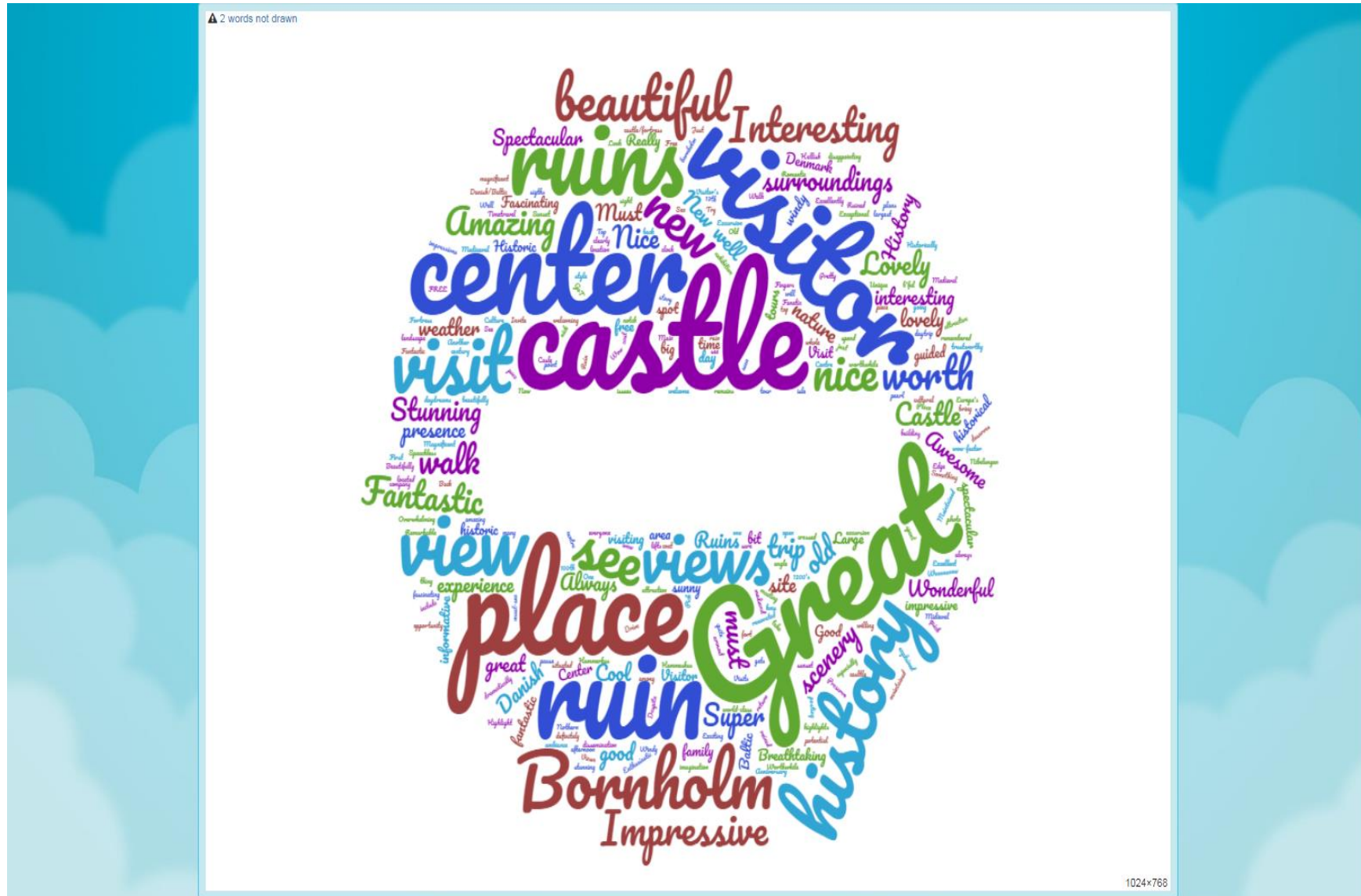
Wordcloud, Visby City Wall, Tripadvisor

Thank you for your attention !