# English-language social media in Finland: Twitter data collection and analysis

Steven Coats, University of Oulu

31 August 2014 – 12th ESSSE Conference, SLANG27



# Outline

1. Twitter as a ”Linguistic Landscape”
2. Methods
   - Compiling the Finland tweet data, processing the tweets, excluding non-English tweets, preparing the comparison corpus
3. Preliminary analysis
   - Rank/frequency profiles for lexical items, lexical clusters, collocations/n-grams
4. Preliminary conclusions



# Twitter as a Linguistic Landscape?

- Linguistic landscapes typically describe the use of language in public physical spaces and thus provide insight into localized issues of sociolinguistic interest such as language use and identity, particularly in bi- and multilingual environments (Landry and Bourhis 1997, Shohamy and Gorter 2008)
- Linguistic landscapes are real physical spaces?



# Twitter as a linguistic landscape?

BUT:

- ”It has to be determined what belongs to the linguistic landscape.” (Gorter 2006: 3)
- ”Almost all humans today live in a textually mediated world, and the texts which mediate and impact on our lives are by no means all fixed in (physical) space.” (Sebba 2010: 61)
- Online and virtual space constitute increasingly important domains of language use for much of the world’s population.
- An expanded linguistic landscape concept could incorporate online media
- Of particular interest would be media platforms that represent sites of bi- and multilingual interaction
- Twitter: ~500m users, ~340m tweets per day



# Some recent Twitter research

| Field | Nr. articles |
|---|---|
| Law | 2 |
| Physics | 2 |
| Mathematics | 3 |
| Geography | 4 |
| Sociology | 10 |
| Political Science | 12 |
| Medicine | 15 |
| Economics | 29 |
| Information science (including some language/linguistics studies) | 37 |
| Communications | 50 |
| Computer science | 64 |

(Boyd 2014)



# Some recent Twitter linguistics/NLP research

- Developing accurate Twitter-English translation and PoS tagging tools (Gimpel et al. 2011; main problem is the high degree of orthographic variation)
- Emoticon sociolinguistics :--DD :) (-;
  - Tweet lexical/emoticon frequencies and demographic variables e.g. gender (Bamman, Eisenstein and Schnoebelen 2014, Schnoebelen 2012)
- Twitter geographical variation/dialectology (Eisenstein et al. 2010)

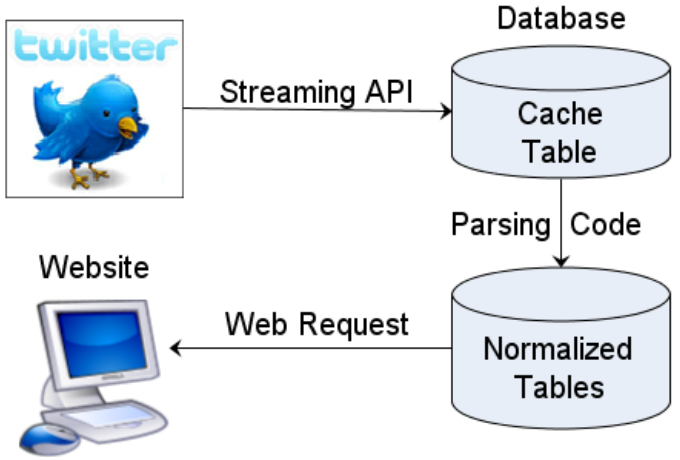
# This project: steps towards characterizing an emerging online English variety?

- To what extent are people in Finland using English in Twitter?
- Is English-language Twitter in Finland different from "normal" Twitter English? Steps:
  - Compile corpus of English-language tweets originating from Finland
  - Compare with corpus with no geographical restrictions



# Compiling the data: Twitter API

- Better than native Twitter web interface (e.g. number, chronological range, geocoding, other features of accessible Tweets)
- Unlimited Twitter stream ("firehose") is proprietary big data, only available to companies working in the "Twitter ecosystem"
- 1% stream available to all



# Tweepy

- Interact with the Twitter API using Python
- Automate authentication process
- Customizable

# Collecting tweets: Python script

twexpGeoFin1.py - C:\Python27\TweetData\twexpGeoFin1.py

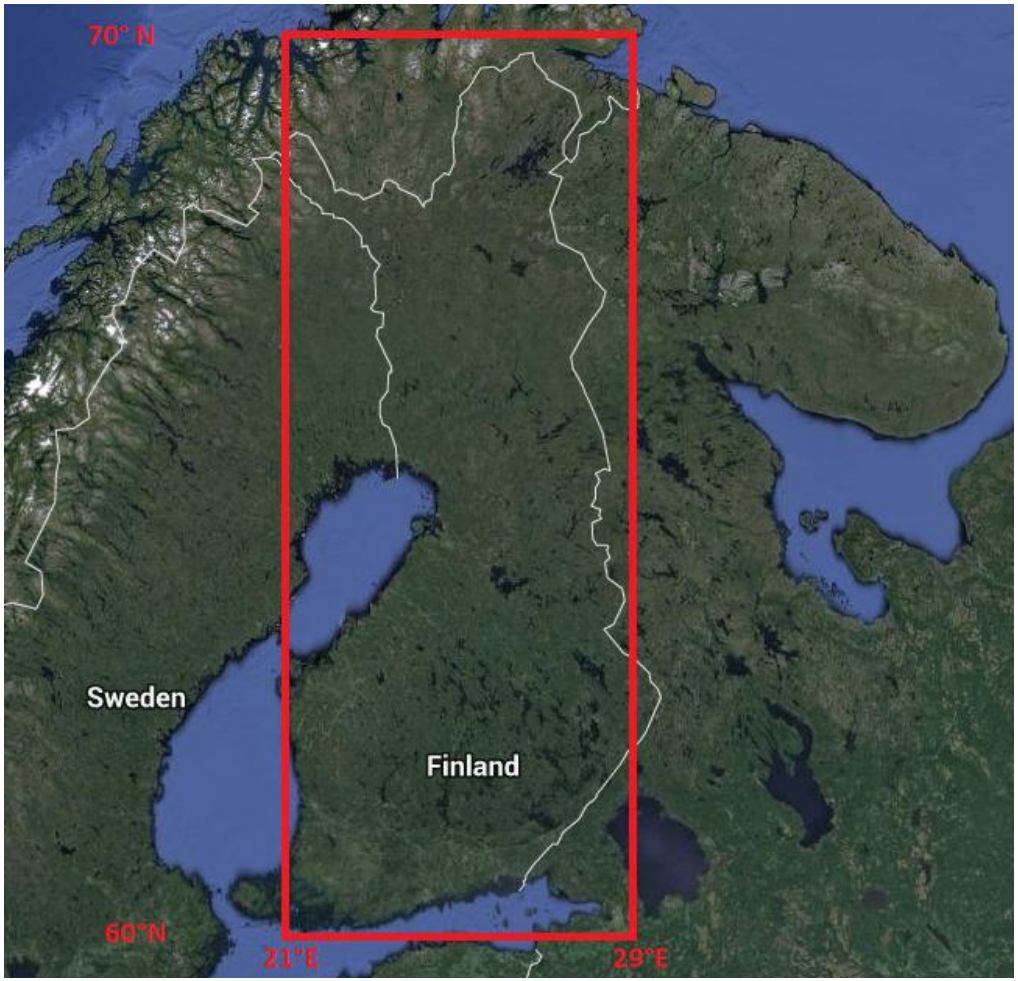
File Edit Format Run Options Windows Help

```
# -*- coding: utf-8 -*-
import sys
import tweepy
import unicodedata
import codecs

consumer_key="yiszUZUdmISP7AiagyLHQ"
consumer_secret="w2InqQjQCSO9zC3keSodjC0yoe7lkNGKQF5ZtbtveMM"
access_key = "262579125-h3ZiXGRm2b3aZuNFyly5YbYStuxwkxWpRoLozLy7"
access_secret = "jWoKh50GOz2OZ7KOTjpjWNdhaF4y25Jrt6S6tKA"


auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_key, access_secret)
api = tweepy.API(auth)

class CustomStreamListener(tweepy.StreamListener):
    def on_status(self, status):
        try:
            print "%s\n%s\n%s\n%s\n%s\n\n" % (status.text,
                                    status.author.screen_name,
                                    status.created_at,
                                    status.source,status.coordinates,status.place)
            with codecs.open('test1.txt', 'ab','utf-8') as f:

                        newline=' NEWLINE'
                        linebreak='\r\n'
                        mylist=(status.text, status.author.screen_name, status.created_at, status.source,status.coordinates,status.place)
                        f.write (("%s\t%s\t%s\t%s\t%s\t%s\t") % (mylist)+linebreak)
        except Exception, e:
            print >> sys.stderr, 'Encountered Exception:', e
            pass

    def on_error(self, status_code):
        print >> sys.stderr, 'Encountered error with status code:', status_code
        return True # Don't kill the stream

    def on_timeout(self):
        print >> sys.stderr, 'Timeout...'
        return True # Don't kill the stream

sapi = tweepy.streaming.Stream(auth, CustomStreamListener())
sapi.filter(locations=[21,60,29,70])
```

Ln: 1 Col: 0

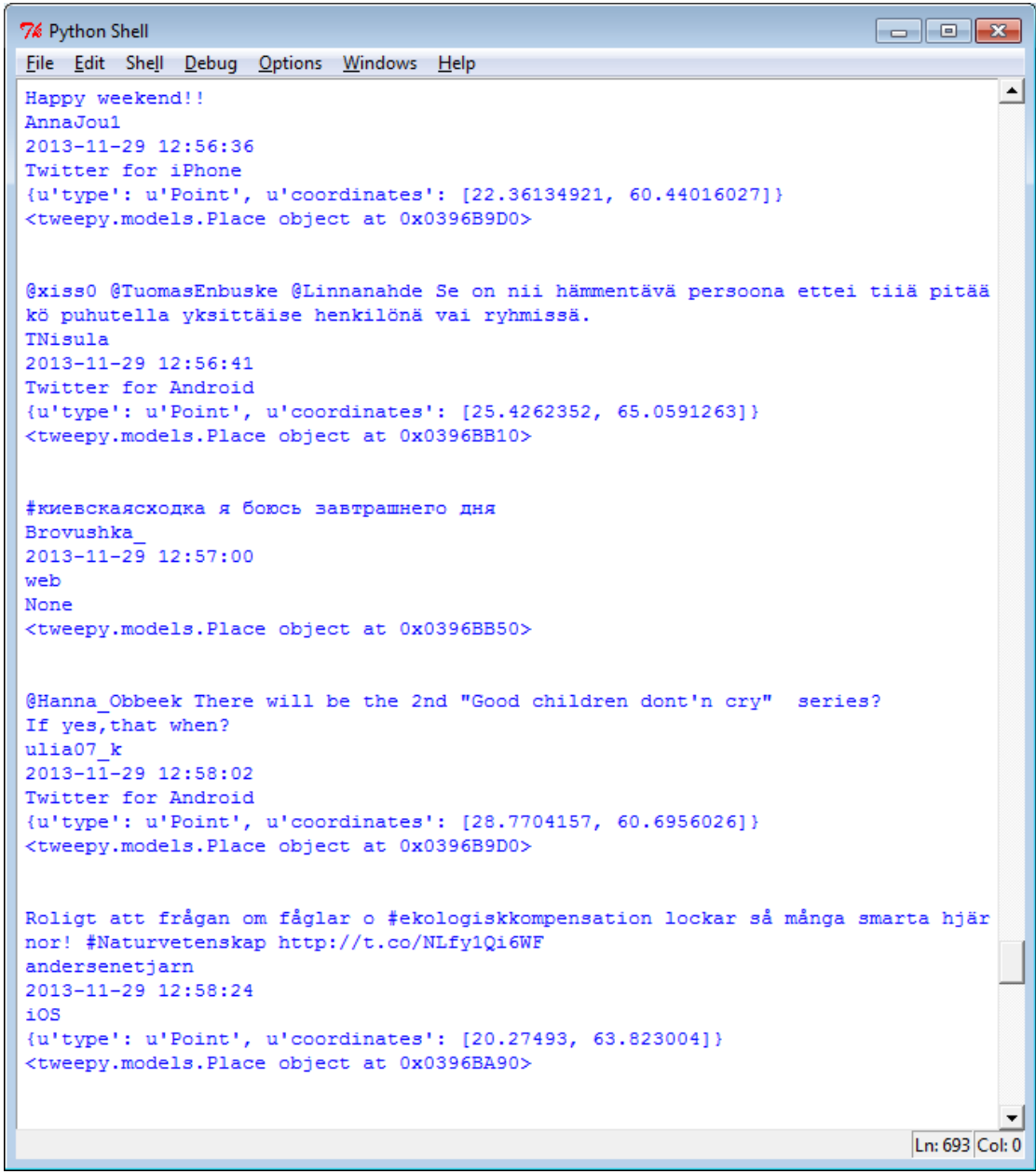
# Limiting geographical range of collected tweets

- Most tweets do not have geocoordinates, users can opt-in
- Summer 2013: Geolocation protocol changed in API (not just long/lat)
- Twitter stream access: 1% of all available tweets returned
- Tweets from Sweden, Norway and Russia returned as well

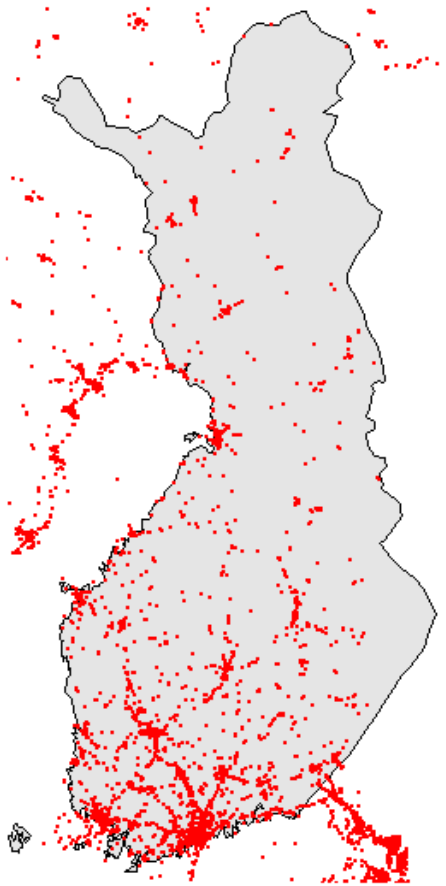


# Output

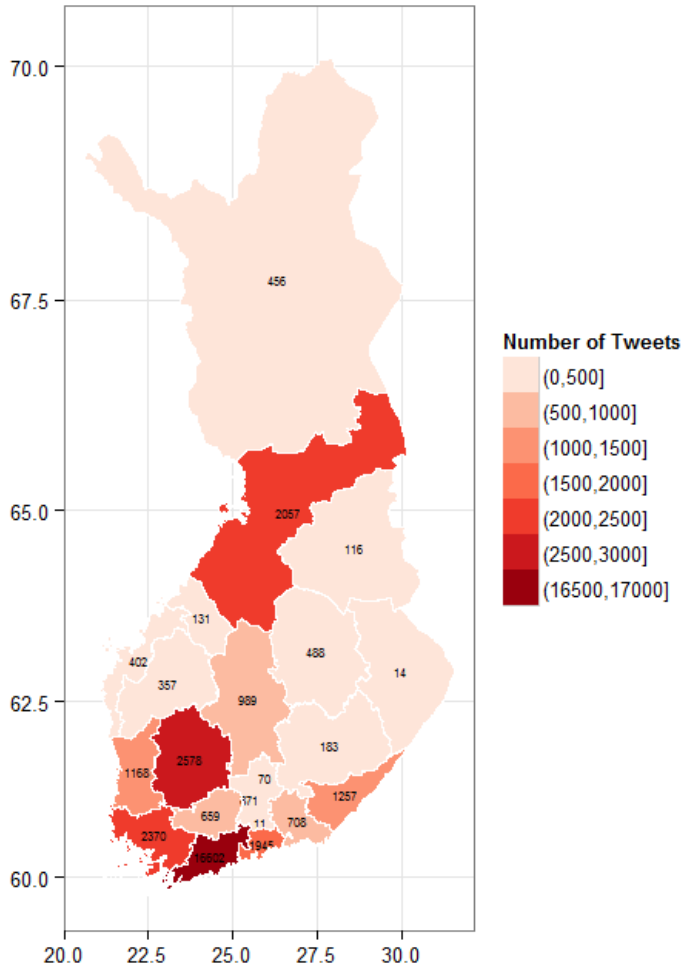
- ~139,000 tweets
- ~1,152,000 words
- Tagged with geographic location (longitude and latitude)
- Finland English Corpus



# Geographical distribution of Finland English Corpus material



Location of individual tweets harvested by script



# Output: Comparison Corpus

- Global tweet database compiled 2008-9 at Texas A&M University, Texas, USA
- ~299,000 tweets
- ~3,610,000 words
- No geographical tags
- Comparison Corpus

```
214352  @halflite Deborah Coxの追加しました．ついでに"Dub Be Good to Me"なんてのも． http:/
214353  finetuneにはSOS Bandの曲がほとんどないみたい．代わりに他のアーティストによる関連曲を2曲入れた
214354  プリンスとジャム＆ルイス，どちらがテクノロジーをうまくつかってきたか？使い方の違いは？とか．
214355  あらためて考えてみれば，おもしろい論点がいろいろあるんだよな．NJSの前の時期ってことにもなるわけだし．バンドと(
214356  95年のBMRに，80年代ファンク特集がある．創刊200号特集で，4号連続で掲載．
214357  あ，45曲か
214358  ジャムルイハ十年代ものだけでやりたかったけど，アーティスト一人3曲しばりで40曲もとてもしかんかったので，九十年
214359  ゆうきさんのを見て，finetuneやってみる．ジャム＆ルイス関連で選曲＝http://www.finetune.com/
214360  blog付属の小さな掲示板みたく使えるなら使ってみたい
214361  I'm happy about two announcements: 1. Obama announces today Hillary i
214362  I'm having coffee with poet Dan Albergotti, watching the rain fall on
214363  I'm heading up to Clemson with Dan Albergotti. Check out his recent b
214364  I'm fighting the drowsy effects of tryptophan and watching the terror
214365  LIke Ben Franklin, I think beer is proof that God loves us and wants
214366  Check out poet Dan Albergotti: The Boatloads, BOA Editions, 2008; and
214367  I'm making weekend plans with poet Dan Albergotti. Check out his rece
214368  Ain't walkin'. Just talkin'. Walkin' through this weary world of woe.
214369  rainy day women must wait.
214370  I'm revisiting "The Tale of Sweeney Todd" with Ben Kingsley for the s
214371  Laughing at AIG exec. Ed. Libby's agreement to slash his salary to a
214372  I'm thinking life in Bangkok must be like life elsewhere: rough.
214373  I'm studying the rimes of frost.
214374  I might be the only Clemson Tiger fan on twitter, but I can't wait fo
214375  @JRPoole I love old Soundgarden.
214376  teaching James Joyce's "Araby" and trying, between classes, to catch
214377  teaching the virtues of Robert Frost, William Blake, and Mark Twain.
```



# Cleaning up the tweets: regex

- Removing #hashtags (including #rt retweets), @usernames, urls:
  - ((mailto\:|(news|(ht|f)tp(s?))\://){1}\S+) ; @\\w+; #\\w+
- Identifying and removing (some) automated tweets of limited linguistic interest
  - Automated weather reports, automated hourly tweets announcing the time, e.g.



# Processing: Language detection

- English, Finnish, Swedish, Russian, Norwegian, Danish, French, Chinese, German, Polish, Japanese, (etc.) tweets
- Language detection: chromium compact language detection, a Python binding of chromium (McCandless/Sites 2013)
- Utilizes probability matrices for character 4-grams to assign language
- Each individual tweet was tagged with the most likely language according to the assignation algorithm from the cld script



# Language distribution

| Comparison Corpus | | | Finland English Corpus | | |
|---|---|---|---|---|---|
| | Number of tweets | % of total | | Number of tweets | % of total |
| English | 182244 | 60.9 | English | 41497 | 29.8 |
| Finnish | 59 | 0.01 | Finnish | 27977 | 20.1 |
| Swedish | 138 | 0.05 | Swedish | 10525 | 7.6 |
| Russian | 952 | 0.3 | Russian | 8906 | 6.4 |
| French | 924 | 0.3 | French | 1221 | 0.9 |
| Norwegian/ Danish | 213 | 0.07 | Norwegian /Danish | 1208 | 0.9 |
| Turkish | 318 | 0.1 | Turkish | 600 | 0.4 |
| Japanese | 31058 | 10.4 | Japanese | 344 | 0.2 |
| Mandarin | 11580 | 3.9 | Mandarin | 299 | 0.2 |
| German | 1777 | 0.59 | German | 165 | 0.1 |
| Spanish | 9652 | 3.2 | Arabic | 76 | 0.05 |
| Unknown | 40907 | 13.7 | Unknown | 40270 | 28.9 |
| Others | 19620 | 6.6 | Others | 5938 | 4.3 |
| Σ | 299442 | 100 | Σ | 139008 | 100 |

# Corpora basic statistics, after processing steps

## Finland English Corpus

```
Sample size:      N  = 362351
Vocabulary size: V  = 32934
Range of freq's: f  = 1 ... 11775
Mean / median:   mu = 11.00234 ,  M = 1
Hapaxes etc.:    V1 = 20735 ,  V2 = 4137
```

## Comparison Corpus

```
Sample size:      N  = 2536815
Vocabulary size: V  = 306332
Range of freq's: f  = 1 ... 77829
Mean / median:   mu = 8.28126 ,  M = 1
Hapaxes etc.:    V1 = 255988 ,  V2 = 14991
```



# Analysis: Rank/frequency profiles and vocabulary growth measures

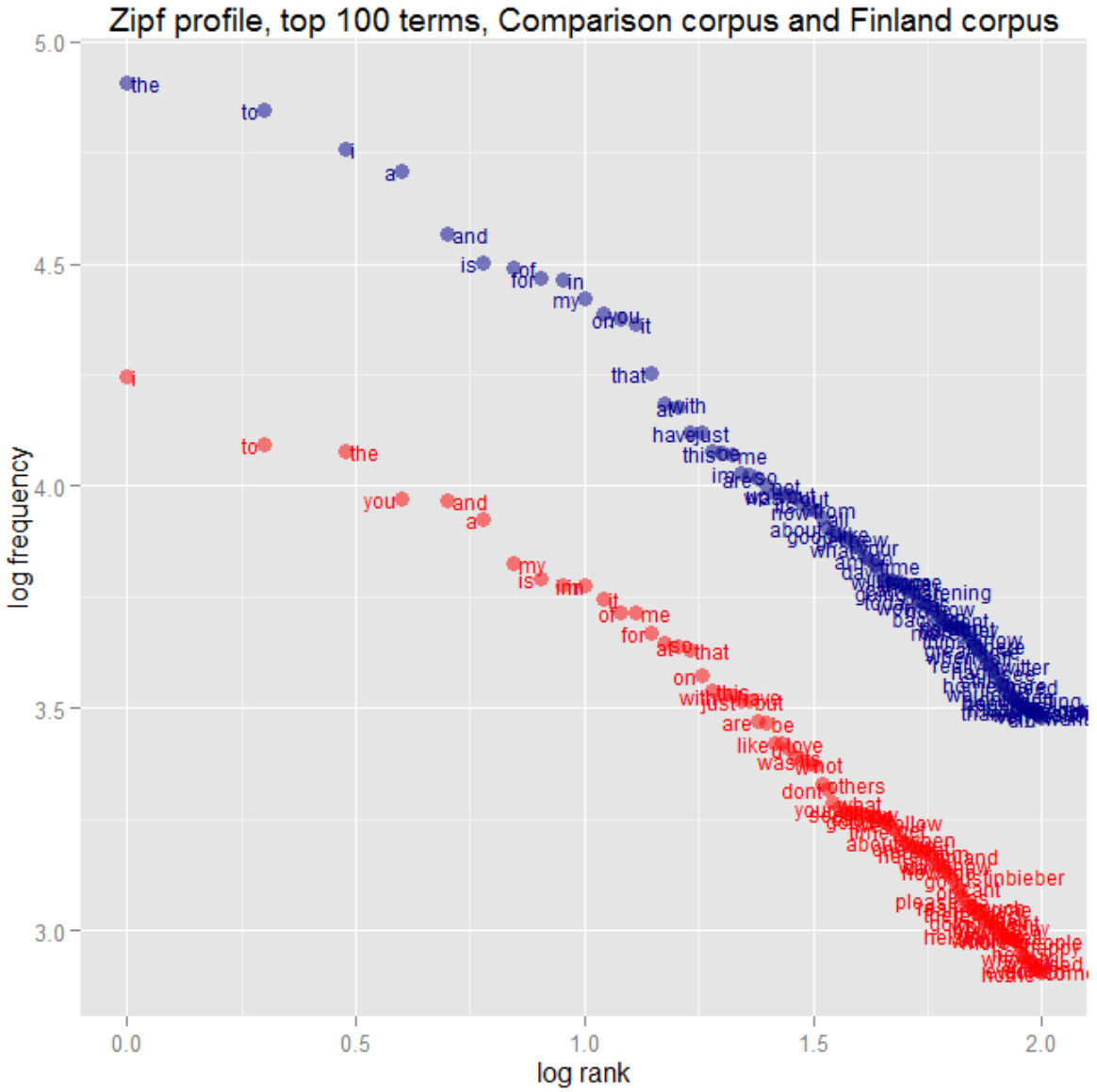
| k | f | type |
|---|---|---|
| 1 | 11777 | i |
| 2 | 8241 | to |
| 3 | 8017 | the |
| 4 | 6261 | and |
| 5 | 5664 | a |
| 6 | 5213 | you |
| 7 | 4622 | my |
| 8 | 4146 | im |
| 9 | 4080 | in |
| 10 | 3995 | is |
| 11 | 3598 | of |
| 12 | 3569 | it |
| 13 | 3149 | for |
| 14 | 3116 | me |
| 15 | 3084 | at |
| 16 | 2913 | so |
| 17 | 2812 | that |
| 18 | 2356 | on |
| 19 | 2325 | this |
| 20 | 2249 | just |

| k | f | type |
|---|---|---|
| 1 | 80282 | the |
| 2 | 70000 | to |
| 3 | 57199 | i |
| 4 | 51071 | a |
| 5 | 36912 | and |
| 6 | 31890 | is |
| 7 | 31063 | of |
| 8 | 29301 | for |
| 9 | 28979 | in |
| 10 | 26432 | my |
| 11 | 24476 | you |
| 12 | 23840 | on |
| 13 | 23145 | it |
| 14 | 17967 | that |
| 15 | 15292 | with |
| 16 | 15035 | at |
| 17 | 13220 | just |
| 18 | 13151 | have |
| 19 | 11978 | be |
| 20 | 11872 | this |

20 most frequent terms, Finland English Corpus and Comparison Corpus



- Rank-frequency profile exhibits familiar Zipfian shape
- What about terms that occur infrequently?



# Grouped frequency distribution or frequency spectrum

- *Hapax legomenon*:

a word that occurs once in a text/corpus

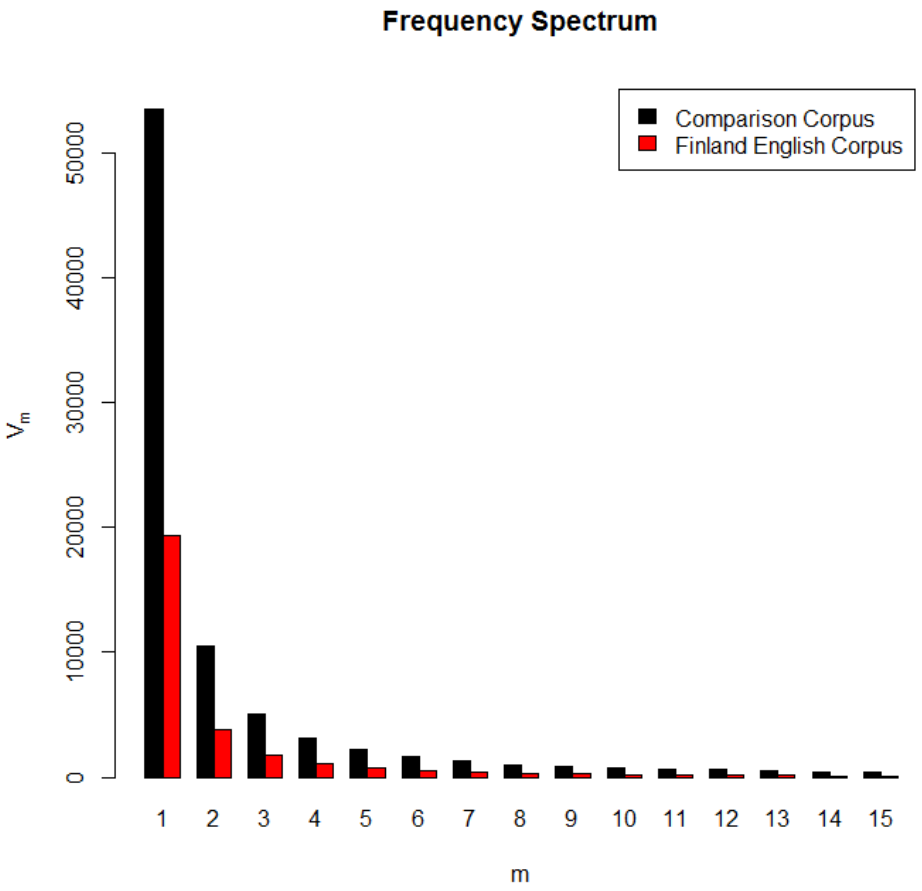
- *Dis legomenon:*

a word that occurs twice in a text/corpus

- Etc.

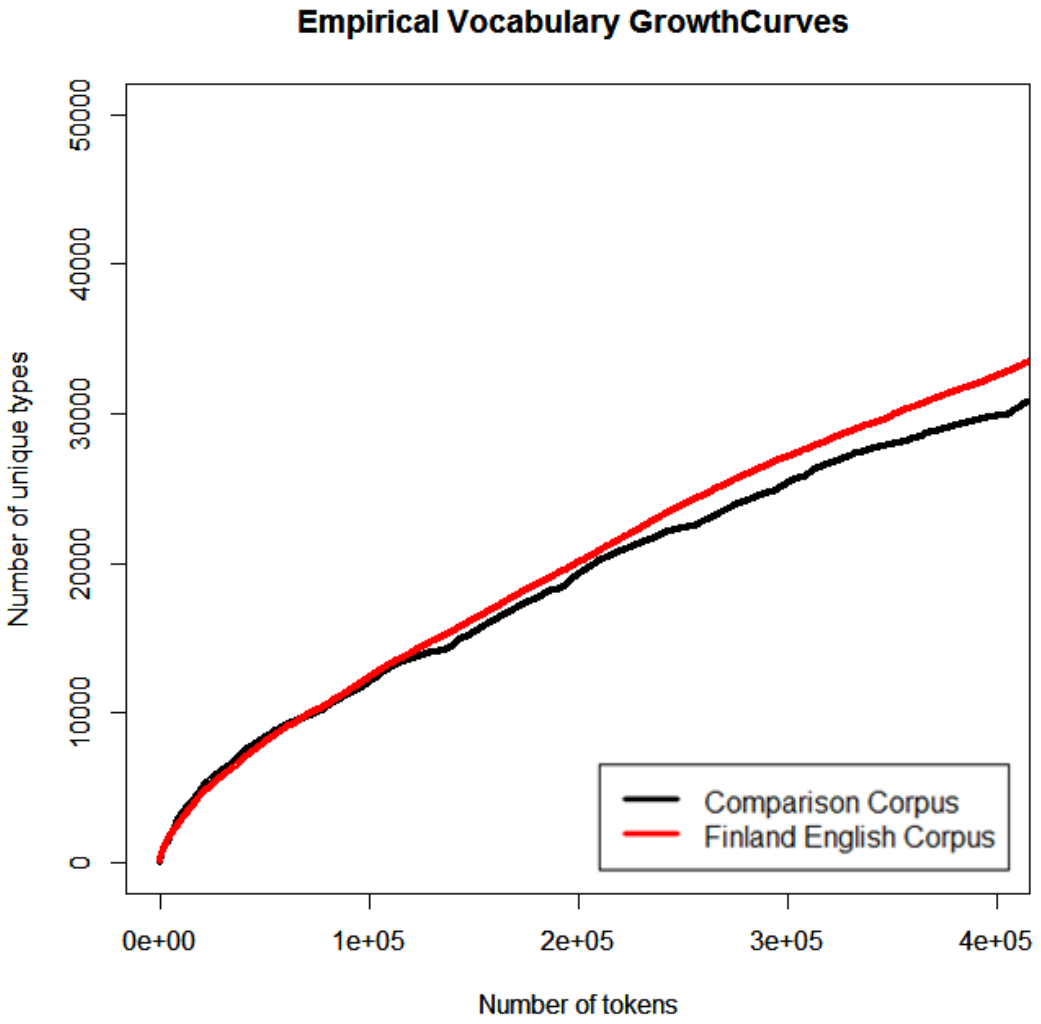
$$V(m,N) = \sum_{i=1}^{V(N)} I_{[f(i,N)=m]} :$$

the number of types with frequency $m$ in a sample of $N$ tokens (Baayen 2001: 8)



# Vocabulary growth

- Type-token ratio and vocabulary growth rate (rate at which new words are added to the corpus) can indicate lexical richness
- V(1,N)/V of Finland English Corpus grows somewhat more slowly than does that of Comparison Corpus initially, but then increases
- But: measures such as type-token ratio are strongly influenced by corpus size (Baayen 2001: 24ff.)

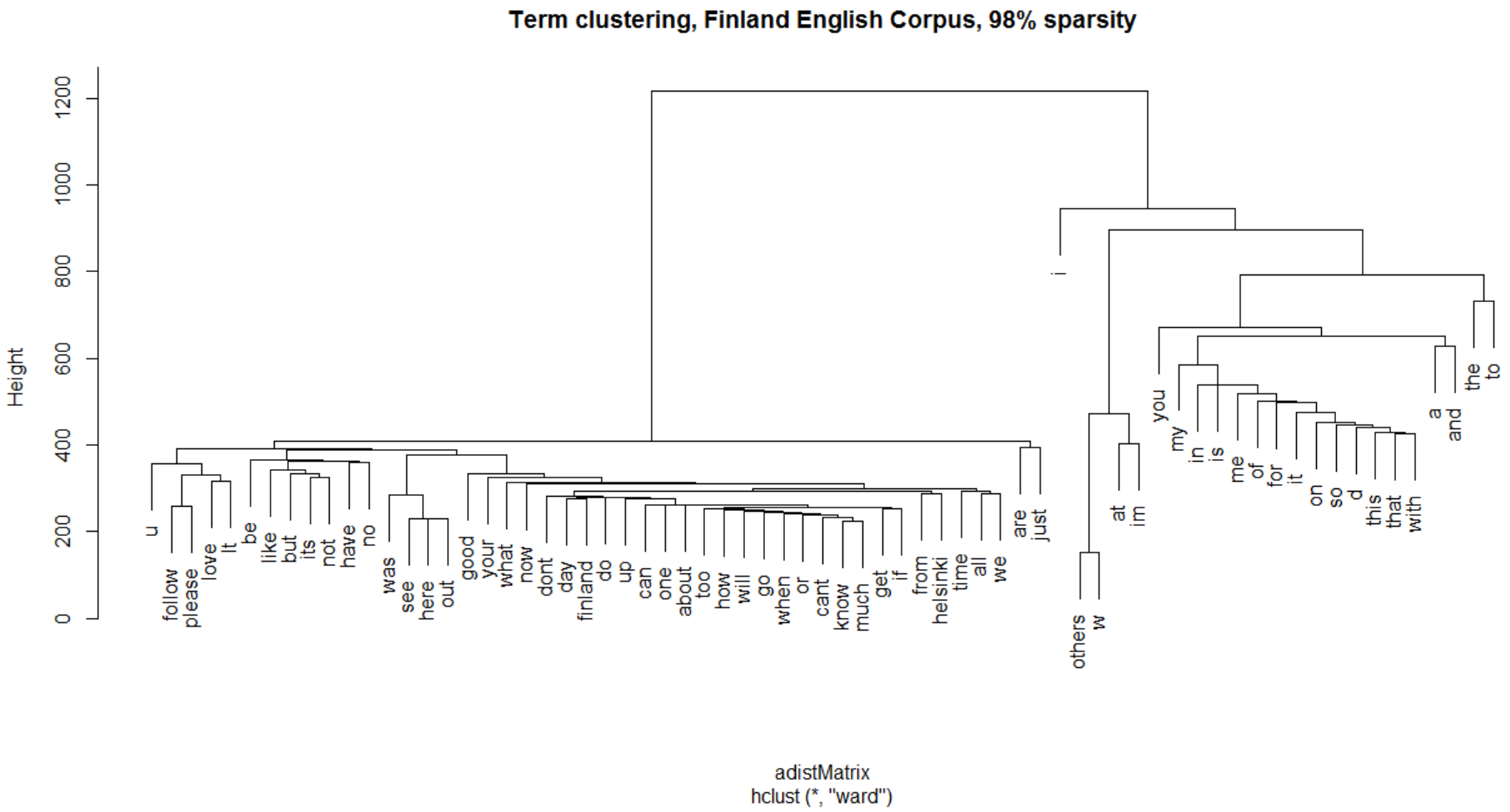


# Lexicon: clustering

- Clustering shows which words are likely to appear together in a tweet
- A matrix is created in which axes are all word types and all individual tweets.
- Using least squares method (Ward's), determine geometric distance between all matrix elements.
- Remove word types that have extremely low frequencies
- R text mining package (tm)



# Lexicon: clustering algorithms



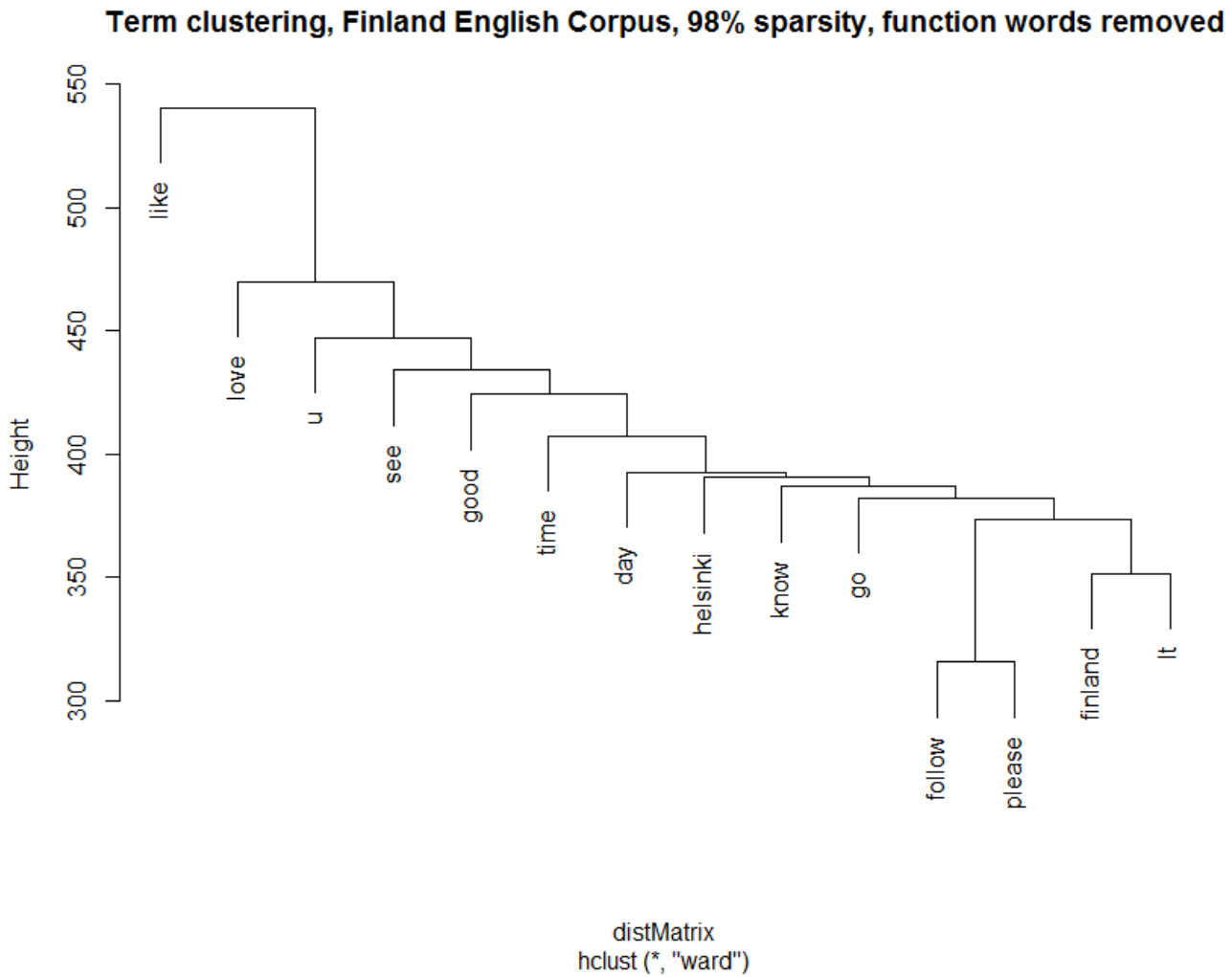
# Lexicon: clustering algorithms

- Remove common English function words, including articles, conjunctions, pronouns, demonstratives, numbers, modal verbs

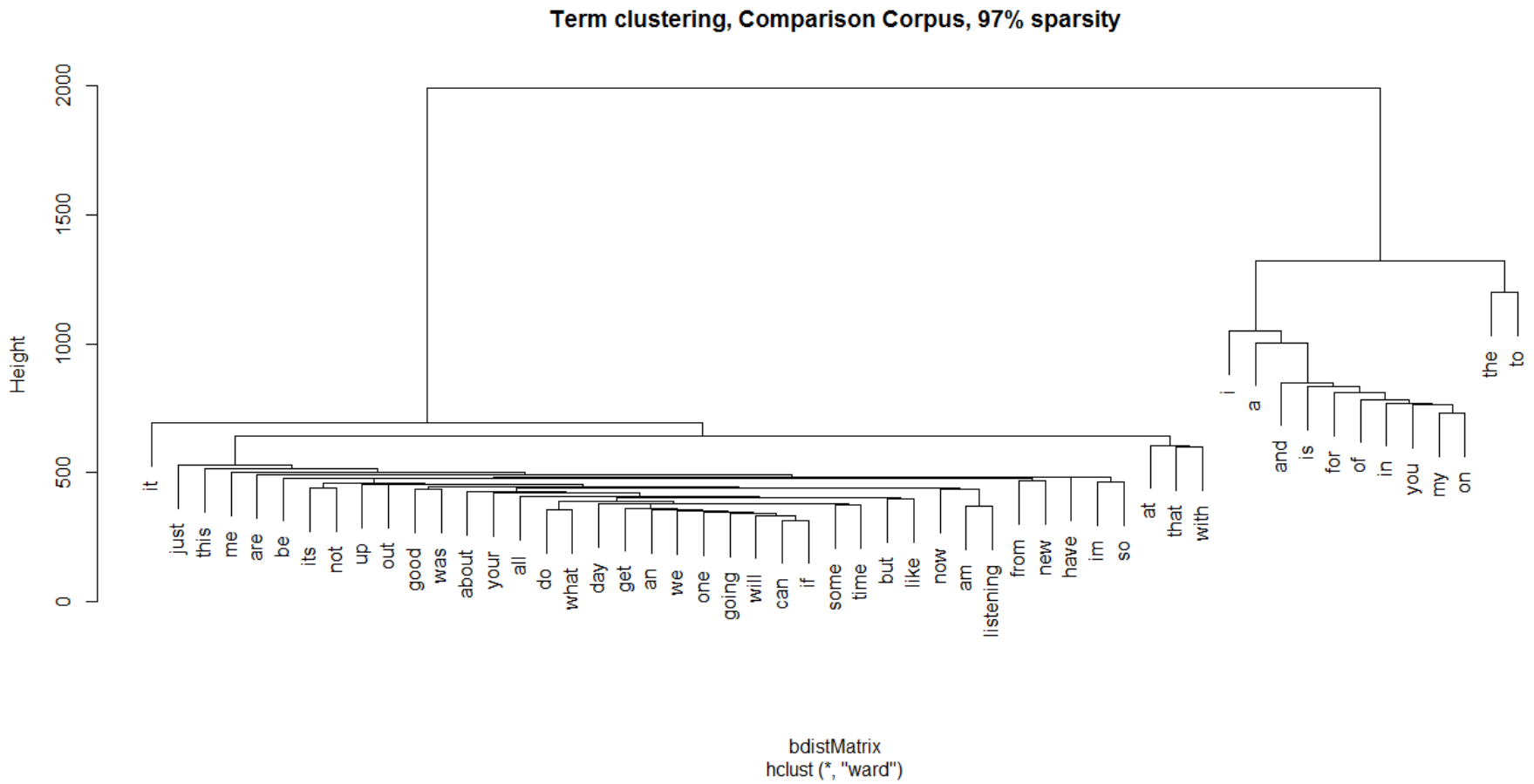
```
 [1] "a"          "about"      "above"      "after"      "again"
 [6] "ago"        "all"        "almost"     "along"      "already"
[11] "also"       "although"   "always"     "am"         "among"
[16] "an"         "and"        "another"    "any"        "anybody"
[21] "anything"   "anywhere"   "are"        "arent"      "around"
[26] "as"         "at"         "back"       "else"       "be"
[31] "been"       "before"     "being"      "below"      "beneath"
[36] "beside"     "between"    "beyond"     "billion"    "billionth"
[41] "both"       "each"       "but"        "by"         "can"
[46] "cant"       "could"      "couldnt"    "did"        "didnt"
[51] "do"         "does"       "doesnt"     "doing"      "done"
[56] "dont"       "down"       "during"     "eight"      "eighteen"
[61] "eighteenth" "eighth"     "eightieth"  "eighty"     "either"
[66] "eleven"     "eleventh"   "enough"     "even"       "ever"
[71] "every"      "everybody"  "everyone"   "everything" "everywhere"
[76] "except"     "far"        "few"        "fewer"      "fifteen"
[81] "fifteenth"  "fifth"      "fiftieth"   "fifty"      "first"
[86] "five"       "for"        "fortieth"   "forty"      "four"
[91] "fourteen"   "fourteenth" "fourth"     "hundred"    "from"
[96] "get"        "gets"       "getting"    "got"        "had"
```



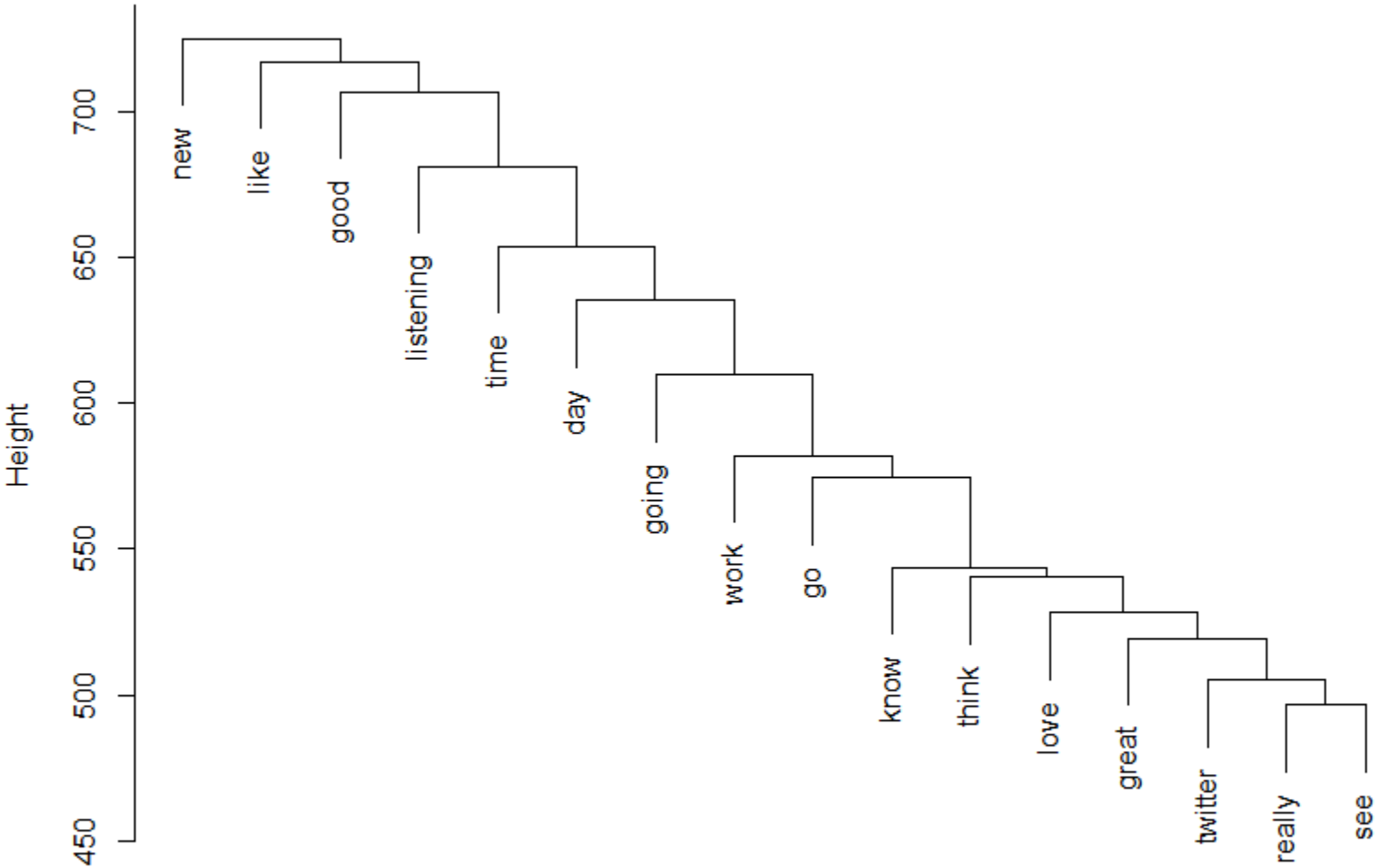
# Lexicon: clustering algorithms



# Lexicon: clustering algorithms



## Term clustering, Comparison Corpus, 98% sparsity, function words removed



# Collocations in the Finland English Corpus: bi- and trigrams

```
1  "('im', 'at')","1499"
2  "('in', 'the')","688"
3  "('i', 'have')","677"
4  "('i', 'dont')","664"
5  "('i', 'love')","630"
6  "('follow', 'me')","558"
7  "('i', 'just')","546"
8  "('to', 'be')","526"
9  "('and', 'i')","519"
10 "('going', 'to')","492"
11 "('helsinki', 'w')","477"
12 "('of', 'the')","473"
13 "('2', 'others')","442"
14 "('i', 'cant')","431"
15 "('but', 'i')","416"
16 "('this', 'is')","404"
17 "('for', 'the')","388"
18 "('on', 'the')","387"
19 "('i', 'am')","379"
20 "('i', 'was')","372"
```

```
1  "('w', '2', 'others')","441"
2  "('w', '3', 'others')","255"
3  "('i', 'love', 'you')","210"
4  "('was', 'out', 'running')","175"
5  "('i', 'want', 'to')","162"
6  "('w', '4', 'others')","147"
7  "('helsinki', 'w', '2')","145"
8  "('i', 'have', 'to')","143"
9  "('hel', 'vantaa', 'w')","134"
10 "('at', 'hel', 'vantaa')","133"
11 "('im', 'at', 'hel')","133"
12 "('im', 'going', 'to')","120"
13 "('i', 'dont', 'know')","116"
14 "('im', 'at', 'kotilinnake')","115"
15 "('w', '5', 'others')","101"
16 "('follow', 'me', 'i')","100"
17 "('come', 'to', 'finland')","99"
18 "('others', 'im', 'at')","95"
19 "('helsinki', 'w', '3')","89"
20 "('i', 'need', 'to')","87"
```

# Collocations in the Finland English Corpus: 4-grams

```
"('helsinki', 'w', '2', 'others')","145"
"('at', 'hel', 'vantaa', 'w')","133"
"('im', 'at', 'hel', 'vantaa')","133"
"('helsinki', 'w', '3', 'others')","89"
"('as', 'the', 'mayor', 'of')","67"
"('just', 'posted', 'a', 'photo')","67"
"('come', 'to', 'finland', 'ill')","64"
"('finland', 'ill', 'wait', 'for')","64"
"('hope', 'someday', 'youll', 'come')","64"
"('someday', 'youll', 'come', 'to')","64"
"('to', 'finland', 'ill', 'wait')","64"
"('youll', 'come', 'to', 'finland')","64"
"('love', 'you', 'so', 'much')","55"
"('follow', 'me', 'i', 'need')","54"
"('i', 'love', 'you', 'so')","53"
"('w', '2', 'others', 'pic')","52"
"('helsinki', 'w', '4', 'others')","51"
"('i', 'dont', 'want', 'to')","51"
"('have', 'a', 'great', 'time')","43"
"('hope', 'you', 'have', 'a')","43"
```



# Collocations in the Comparison Corpus: bi- and trigrams

| | |
|---|---|
| 1 | "('in', 'the')","6078" |
| 2 | "('listening', 'to')","5694" |
| 3 | "('i', 'am')","5370" |
| 4 | "('on', 'the')","5350" |
| 5 | "('to', 'the')","5060" |
| 6 | "('of', 'the')","4961" |
| 7 | "('for', 'the')","4684" |
| 8 | "('going', 'to')","3888" |
| 9 | "('to', 'be')","3504" |
| 10 | "('i', 'have')","3421" |
| 11 | "('at', 'the')","3058" |
| 12 | "('for', 'a')","2865" |
| 13 | "('i', 'think')","2599" |
| 14 | "('to', 'get')","2512" |
| 15 | "('have', 'a')","2185" |
| 16 | "('is', 'a')","2153" |
| 17 | "('have', 'to')","2108" |
| 18 | "('with', 'the')","2073" |
| 19 | "('to', 'go')","2050" |
| 20 | "('in', 'a')","1985" |

| | |
|---|---|
| 1 | "('lastfm', 'listening', 'to')","1298" |
| 2 | "('listening', 'to', 'the')","1260" |
| 3 | "('am', 'listening', 'to')","1149" |
| 4 | "('i', 'am', 'listening')","1148" |
| 5 | "('new', 'blog', 'post')","827" |
| 6 | "('team', 'hyjak', 'playing')","652" |
| 7 | "('going', 'to', 'be')","596" |
| 8 | "('thanks', 'for', 'the')","584" |
| 9 | "('i', 'have', 'a')","583" |
| 10 | "('i', 'have', 'to')","569" |
| 11 | "('to', 'go', 'to')","555" |
| 12 | "('is', 'going', 'to')","545" |
| 13 | "('i', 'want', 'to')","541" |
| 14 | "('i', 'need', 'to')","530" |
| 15 | "('a', 'lot', 'of')","524" |
| 16 | "('im', 'listening', 'to')","503" |
| 17 | "('looking', 'forward', 'to')","499" |
| 18 | "('i', 'think', 'i')","480" |
| 19 | "('im', 'going', 'to')","478" |
| 20 | "('getting', 'ready', 'to')","464" |

# Collocations in the Comparison Corpus: 4-grams

```
"('i', 'am', 'listening', 'to')","1137"
"('heavy', 'traffic', 'on', 'the')","419"
"('posted', 'by', 'espncoms', 'james')","413"
"('by', 'espncoms', 'james', 'walker')","398"
"('team', 'hyjak', 'playing', 'xbox')","367"
"('hyjak', 'playing', 'xbox', '360')","357"
"('fiddling', 'with', 'my', 'blog')","273"
"('with', 'my', 'blog', 'post')","273"
"('is', 'going', 'to', 'be')","208"
"('playing', 'xbox', '360', 'dashboard')","199"
"('between', 'the', 'junctions', 'with')","183"
"('still', 'writing', 'the', 'new')","182"
"('the', 'new', 'blog', 'post')","182"
"('writing', 'the', 'new', 'blog')","182"
"('playing', 'xbox', '360', 'dashboard')","177"
"('am', 'listening', 'to', 'the')","173"
"('for', 'the', 'first', 'time')","173"
"('dashboard', 'watching', 'a', 'video')","159"
"('360', 'dashboard', 'watching', 'a')","154"
"('xbox', '360', 'dashboard', 'watching')","154"
```

# Issues and further directions

- Corpus compilation
  - Automatic language detection script produces many ”Unknowns”
  - Correction for orthographic variation needed
  - Removing ”noise” in terms of automated tweets
- Corpus analysis
  - Characterization of discourse function of n-grams
  - Correlation of geocoded information with various text or non-text variables (n-gram structure, country, province/region)
  - Comparison with other English language corpora (e.g. Davies’ GloWBe)



# Preliminary conclusions

- Finnish Twitter users seem to use English more than any other language
- Lexical richness in English-language tweets in Finland is not significantly different from that of tweets from elsewhere
- English-language Twitter use in Finland reflects Finland-specific topicality in terms of lexis
  - Local communicative orientation despite globalized nature of the medium
- Collocations or lexical bundles seem to be dominated by tweets that may or may not represent human users
- The online linguistic landscape constituted by Twitter in Finland *"may serve important informational and symbolic functions as a marker of the relative power and status of the linguistic communities inhabiting the territory"* (Landry and Bourhis 1997: 23): Language shift?

# Literature


- Baayen, H. 2001. *Word Frequency Distributions.* Dordrecht: Kluwer.
- Bamman, D., J. Eisenstein, T. Schnoebelen. 2014. Gender Identity and Lexical Variation in Social Media. In: *Journal of Sociolinguistics* 18/2.: 135-160.
- Biber, D. 2009. Multi-dimensional approaches. In: Lüdeling, A., and M. Kytö, eds., *Corpus Linguistics: An International Handbook, V.2.* Berlin: De Gruyter, 2009, 822-854.
- Boyd, D. (2014) Bibliography of research on twitter & microblogging: (www.danah.org/researchBibs/twitter.php)
- Eisenstein, J., B. O'Connor, N. Smith, E. Xing. 2010. A latent variable model for geographic lexical variation. In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1277-1287. (http://dl.acm.org/citation.cfm?id=1870782)
- Evert, S. and Baroni, M. 2007. ZipfR: Word frequency distributions in R. In: *Proceedings of the 45th Annual Meeting of the Association for Computational* Linguistics: 29-32. Prague, Czech Republic. (http://zipfr.r-forge.r-project.org/)
- Feinerer, I., K. Hornik, D. Meyer. 2008. Text Mining Infrastructure in R. In: *Journal of Statistical Software* 25. (http://tm.r-forge.r-project.org/index.html)
- Gimpel, K., N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M Heilman, D. Yogatama, J. Flanigan, N A. Smith. 2011. Part-of-speech tagging for Twitter: annotation, features, and experiments. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*: 42--47.
- Landry, R. and R. Bourhis. 1997. Linguistic Landscape and Ethnolinguistic Vitality: An Empirical Study. In: *Journal of Language and Social Psychology* 16/1: 23-49.
- Manning, C.D, and H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- McCandless, M., Sites, D. 2013. *Compact Language Detector 2*. Computer code. (https://code.google.com/p/cld2/)
- NLTK: http://nltk.org/
- Schneider, E. 2011. *English around the World*. Cambridge, UK: Cambridge University Press.
- Schnoebelen, T. 2012. Do You Smile with Your Nose? Stylistic Variation in Twitter Emoticons. In: *University of Pennsylvania Working Papers in Linguistics* 18/2, Article 14. (http://repository.upenn.edu/pwpl/vol18/iss2/14)
- Tweepy: https://github.com/tweepy/tweepy
- Sebba, M. 2010. Discourses in transit. In: Jaworski, A. and C. Thurlow (eds.), *Semiotic Landscapes: Language, Image, Space,* London: Continuum, 59-76.
- Shohamy, E. and D. Gorter (eds.). 2009. *Linguistic landscape: Expanding the scenery.* New York and London: Routledge.