Repetition Analysis in Lyrics across languages: A case study on English, Dutch and Italian songs

Andrea Favia, Rutger van Deelen

May 2, 2019

Introduction

Psychology

Music as a by product of language development Repetition in music is a cultural universal [Pinker 1997]

About 94% of musical passages are repeated at some point later in the music [Huron and Ollen 2004]

Repetitive music is more enjoyable and draws our attention [Margulis 2013]

Data Mining and Cleaning

Group	Lyrics #
Italian	548
English	14779
OldDutch	220
ModernDutch	437
OldEnglish	304

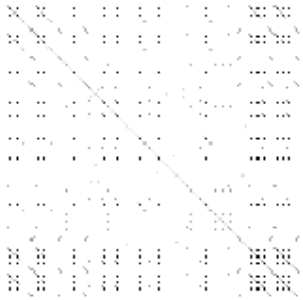
Old lyrics

Old English songs obtained from Archive.org

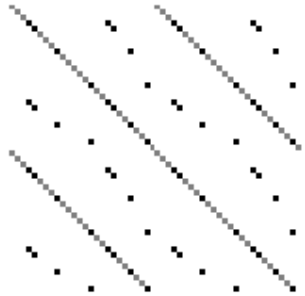
Modern lyrics

Artists selected from different Top 100 lists.
Python scripts to mine the lyrics and process them.
Modern English ones kindly provided by Colin Morris.

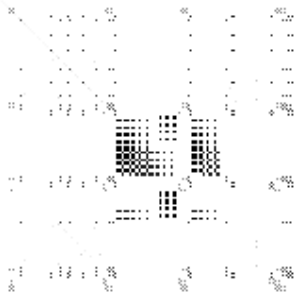
Dot Matrices



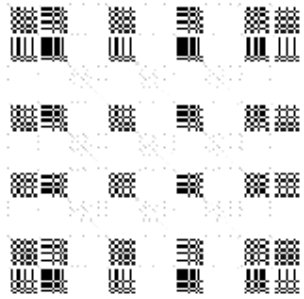
The Killers - Shadowplay



Skrillex - Scary Monsters & Nice Sprites

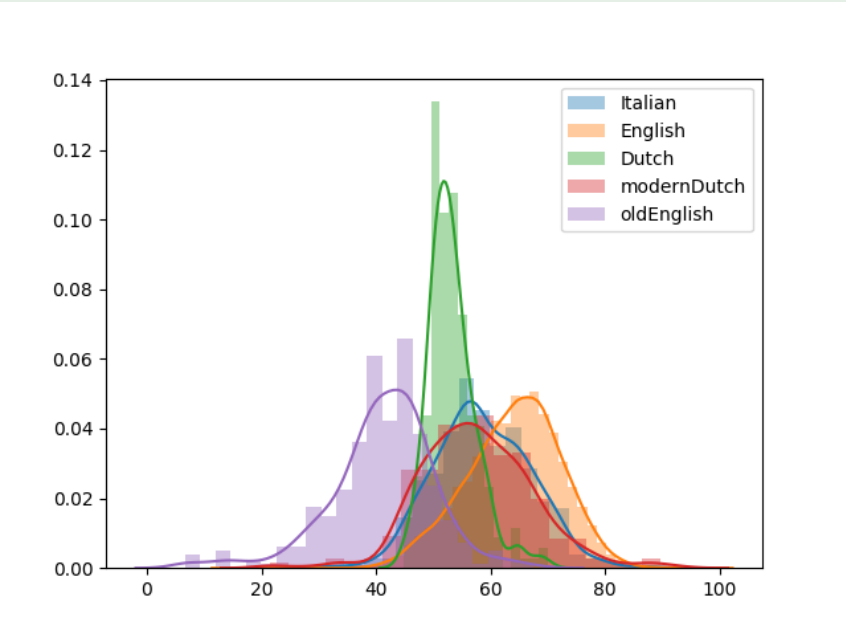


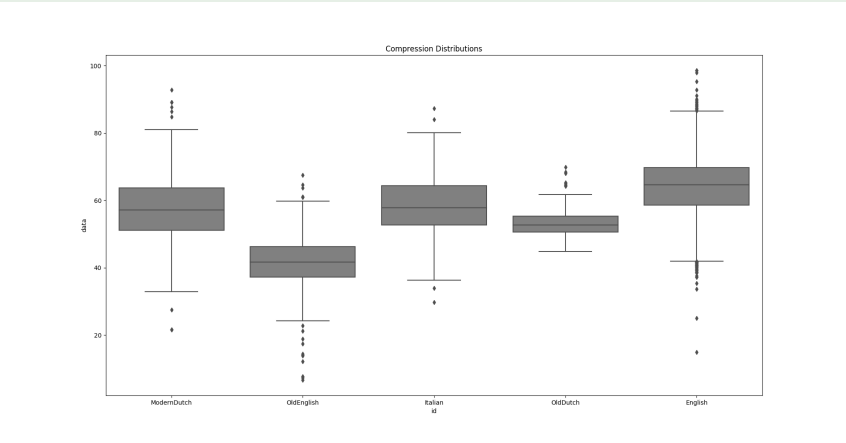
Snap - The Power



The Diamonds - She Say

Compression and Distributions





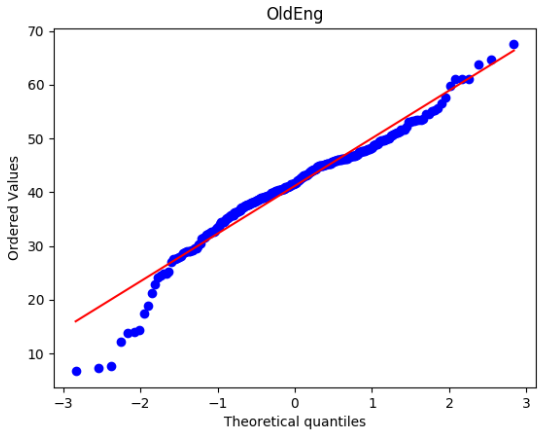
Data processing in 1 simple command

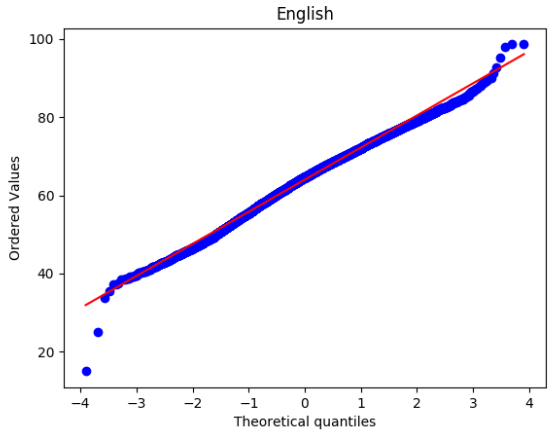
```
gzip -v *txt 2> logfile.log
```

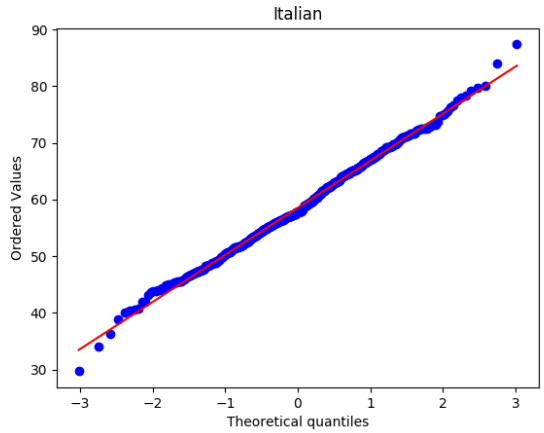
Table: *Shapiro Test results*

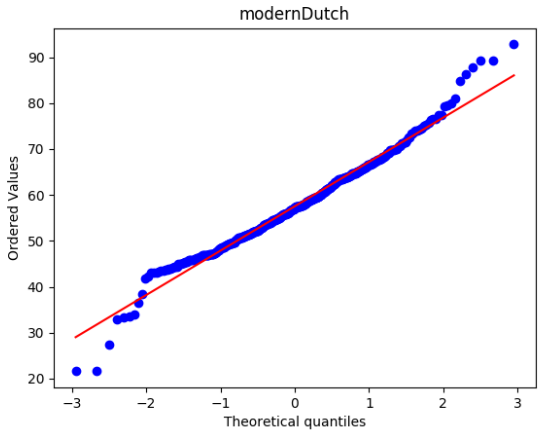
Group	*p-value*
Italian	*0.42*
English	*< .001*
Dutch	*< .001*
modernDutch	*< .001*
OldEng	*< .001*

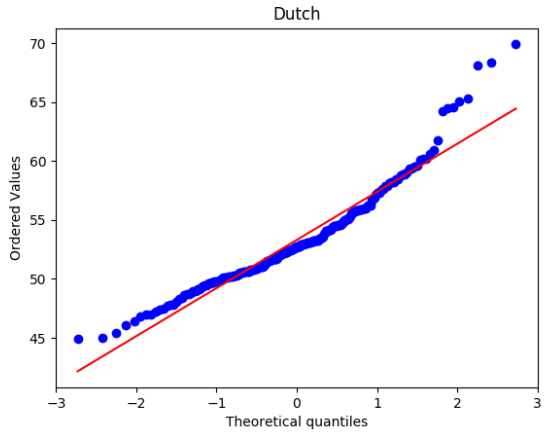
Compression and Distributions











Assumption

Compressibility $\sim=$ Repetitiveness

Difference across time within languages

H1 compressibility is different between time periods within languages

U-Test assumptions:

- ordinal or continuous vars
- indipendent samples
- indipendence of observation

2 Sample K-S assumptions:

- Continuous or ordinal vars
- independent samples
- H_0: Both samples from same set

Table: *U Test*

Groups	*p-value*
DUTCH MODDutch	*< .001*
OLDENG ENG	*< .001*

Table: *KS 2 Sample Test*

Groups	*p-value*
DUTCH MODDutch	*< .001*
OLDENG ENG	*< .001*

Difference across languages

H2 Compressability is different between languages

Table: *U Test*

Groups	*p-value*
ITA ENG	*< .001*
ITA MODDutch	*0.018*
ENG MODDutch	*< .001*

Table: *KS 2 Sample Test*

Groups	*p-value*
ITA ENG	*< .001*
ITA MODDutch	*0.06*
ENG MODDutch	*< .001*

Conclusions

There is difference in lyrics compressibility between languages and between time periods.
There seems to be an incresing trend in compressability from the past until now.

Corpus and Computational linguistics can obtain new valuable insights by approaching things from different perspectives. Interdisciplinarity is a key component in this.

References

David Huron and Joy Ollen. "Musical form and the structure of repetition: A cross-cultural study". In: *Unpublished manuscript* (2004).

Elizabeth Hellmuth Margulis. "Repetition and Emotive Communication in Music Versus Speech". In: *Frontiers in Psychology* 4 (2013), p. 167. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00167. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2013.00167.

Colin Morris. "Are Pop Lyrics Getting More Repetitive?" In: *The Pudding* (2017). URL: https://pudding.cool/2017/05/song-repetition/.

Joseph Nunes, Andrea Ordanini, and Francesca Valsesia. "The Power of Repetition Repetitive Lyrics in a Song Increase Processing Fluency and Drives Market Success". In: *Journal of Consumer Psychology* 25 (Dec. 2014). DOI: 10.1016/j.jcps.2014.12.004.

Steven Pinker. "How does the mind work?" In: *Mind & Language* 20.1 (2005), pp. 1–24. DOI: https://doi.org/10.1111/j.0268-1064.2005.00274.x.

Daniela Sammler et al. "The Relationship of Lyrics and Tunes in the Processing of Unfamiliar Songs: A Functional Magnetic Resonance Adaptation Study". In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30 (Mar. 2010), pp. 3572–8. DOI: 10.1523/JNEUROSCI.2751-09.2010.

Leland Wilkinson and Michael Friendly. "The History of the Cluster Heat Map". In: *The American Statistician* 63.2 (2009), pp. 179–184. DOI: 10.1198/tas.2009.0033. eprint: https://doi.org/10.1198/tas.2009.0033. URL: https://doi.org/10.1198/tas.2009.0033.

R. B. Zajonc. "Attitudinal effects of mere exposure". In: *Journal of Personality and Social*