

The Idea is Staggering

An Introduction to Digital Text Analysis

Hannah Alpert-Abrams
University of Texas at Austin

www.halperta.com | halperta@gmail.com

What is digital textual analysis?

What is digital textual analysis?

Critical Digital Studies

Cultural Analytics

Applications for academic work

What is digital textual analysis?

“... la verdad, cuya madre es la historia, émula del tiempo, depósito de las acciones, testigo de lo pasado, ejemplo y aviso de lo presente, advertencia de lo por venir.”

La historia, *madre* de la verdad; la idea es asombrosa.”

-- Jorge Luis Borges
Pierre Menard, Autor del Quijote

Texts change as they enter new historical contexts.

What is digital textual analysis?

Borges:

“La historia, *madre* de la verdad; la idea es asombrosa.”

Andrew Hurley:

“History, the *mother* of truth!—the idea is staggering.”

Norman Thomas de Giovanni:

“History, the ‘mother’ of truth; the idea is breathtaking.”

James E. Irby:

“History, the mother of truth: the idea is astounding.”

Texts change as they enter new **cultural** contexts.

What is digital textual analysis?

Borges:

“La historia, *madre* de la verdad; la idea es asombrosa.”

Andrew Hurley:

“History, the *mother* of truth!—the idea is staggering.”

Automatic Transcription (OCR):

“history, the mother of truth!-the idea is easierino”

Google Translate:

“History, mother of truth;The idea is amazing.”

Critical digital studies:

Texts change as they enter new digital contexts.

What is digital textual analysis?

Critical Digital Studies

Cultural Analytics

Applications for academic work

What is digital textual analysis?

Word Counts

This corpus has:

15 documents

42,082 total words

7,597 unique word forms

Most frequent words in the corpus:

man (117); time (91); book (62); men (62); secret (52)

created using Voyant Tools
with texts from *Ficciones*

Cultural Analytics:

Digitization enables new ways of analyzing texts.

What is digital textual analysis?

Word Counts

This corpus has:
15 documents
42,082 total words
7,597 unique word forms

Most frequent words in the corpus:
man (117); time (91); book (62); **men** (62); secret (52)



Lemmatization?

Translation?

Masculinity?

created using Voyant Tools
with texts from *Ficciones*

Cultural Analytics:

Digitization enables new ways of analyzing texts.

What is digital textual analysis?

Word Count Analytics (Voyant Tools)

Document Length:

Longest: Tlön Uqbar Orbis Tertius 5,927

Shortest: Theme of the Traitor and Hero 1,211

Vocabulary Density:

Highest: Theme of the Traitor and Hero 0.457

Lowest: Tlön Uqbar Orbis Tertius 0.320

created using Voyant Tools
with texts from *Ficciones*

Cultural Analytics:

Digitization enables new ways of analyzing texts.

What is digital textual analysis?

Word Count Analytics (Voyant Tools)

Document Length:

Longest: Tlön Uqbar Orbis Tertius 5,927

Shortest: Theme of the Traitor and Hero 1,211

Vocabulary Density:

Highest: Theme of the Traitor and Hero 0.457

Lowest: Tlön Uqbar Orbis Tertius 0.320

????

created using Voyant Tools
with texts from *Ficciones*

Cultural Analytics:

Digitization enables new ways of analyzing texts.

What is digital textual analysis?

Topic Modeling:

Each topic is a probability distribution over words

| | |
|---|--|
| 0 | god secret order chance runeberg hladić judas part company act |
| 1 | man told moon labyrinth garden read back hand light ts'ui |
| 2 | dahlmann thought day lonnrot house voice time face felt room |
| 3 | men man time years words night nights place made days |
| 4 | book books library tlön work volume quixote don pages menard |

created using Mallet
with texts from *Ficciones*

Cultural Analytics:

Digitization enables new ways of analyzing texts.

What is digital textual analysis?

Topic Modeling:

Each topic is a probability distribution over words

0 god secret order chance runeberg hladík judas part **company** act

created using Mallet
with texts from *Ficciones*

Cultural Analytics:

Digitization enables new ways of analyzing texts.

What is digital textual analysis?

Topic Modeling:

Each topic is a probability distribution over words

0 god secret order chance runeberg hladiík judas part **company** act

The Company (as it had begun to be called by then) had to protect the interests of the winners, who could not collect their winnings if there was lacking in the coffers the almost entire sum of the fines.

-The Lottery in Babylon

created using Mallet
with texts from *Ficciones*

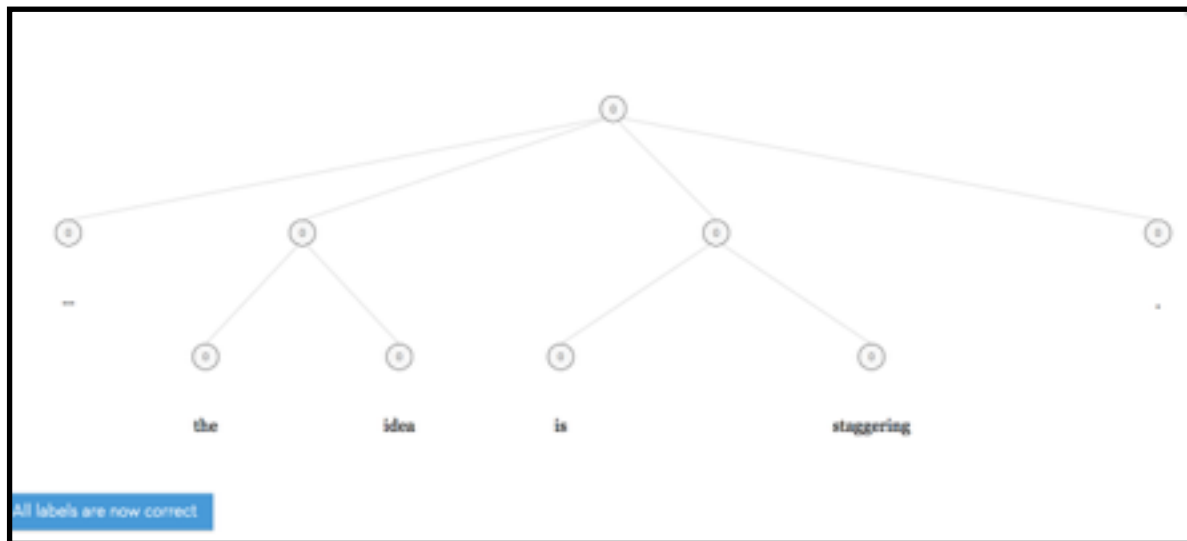
Cultural Analytics:

Digitization enables new ways of analyzing texts.

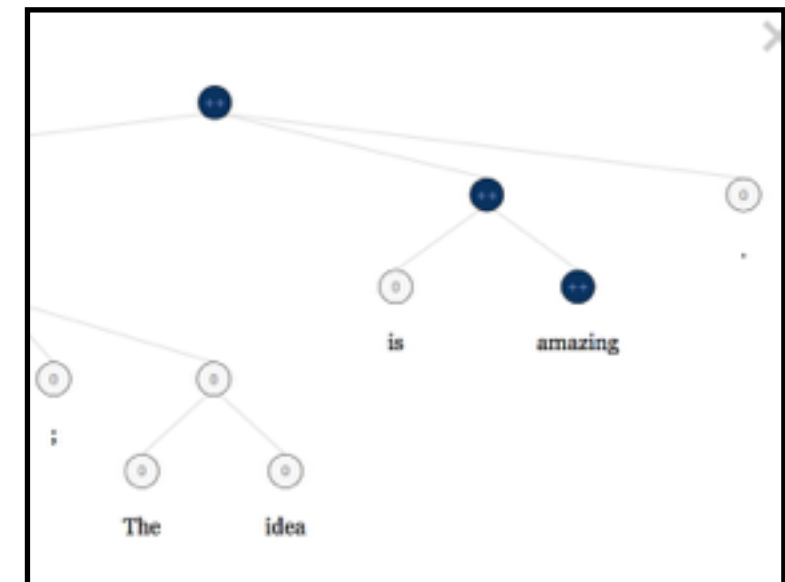
What is digital textual analysis?

<http://nlp.stanford.edu/sentiment/>

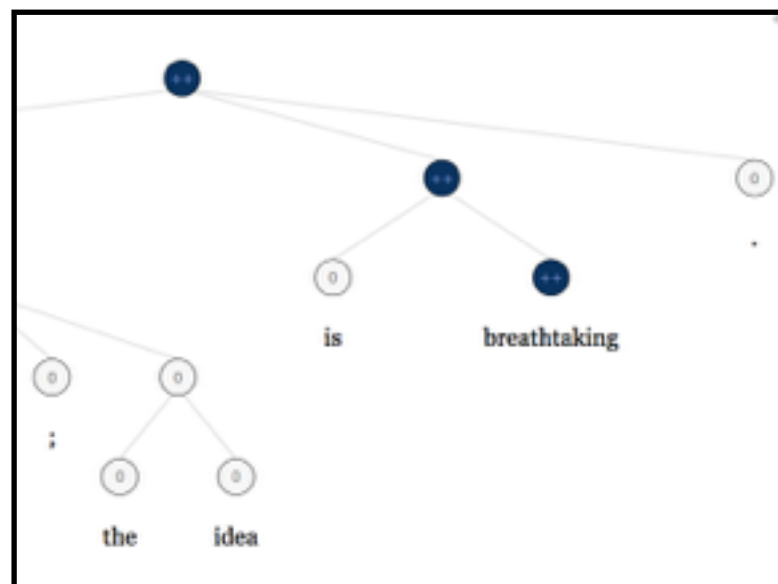
Sentiment Analysis



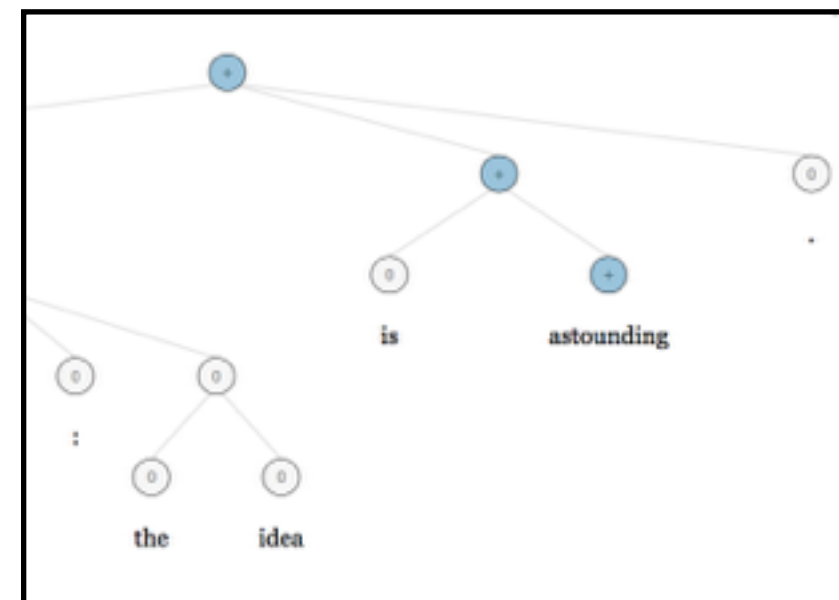
Hurley (Staggering)



Google (Amazing)



Giovanni (Breathtaking)



Irby (Astounding)

Cultural Analytics:
Digitization enables new ways of analyzing texts.

What is digital textual analysis?

Critical Digital Studies

Cultural Analytics

Applications for academic work.

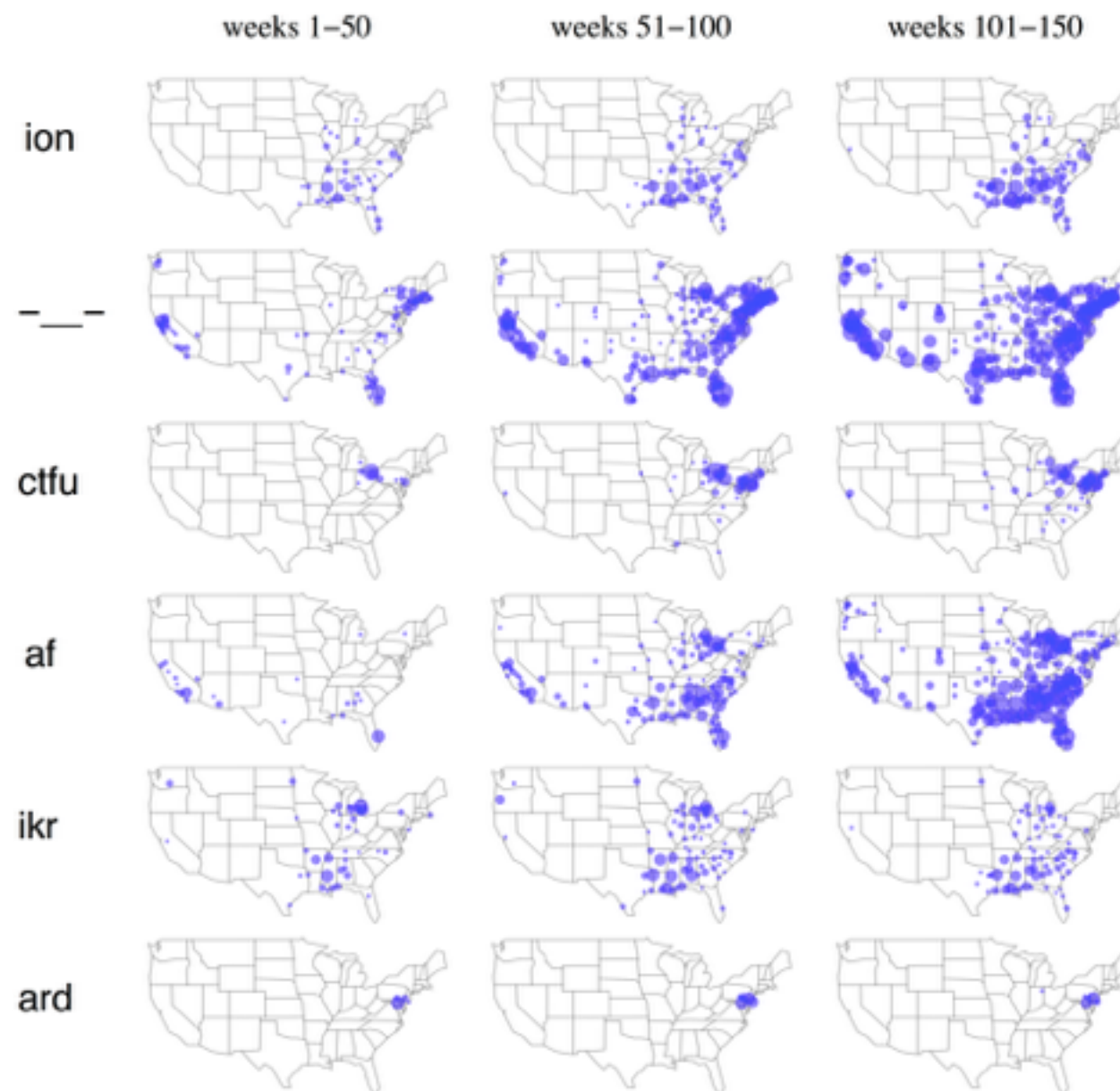
What is digital textual analysis?



Digitizing the 'Sound Explosions' of Anglophone Caribbean Performance Poetry
Janet Neigh

“ By using social media to augment their performance events, the 2 Cents Movement intercepts [the] logic of accumulation by bringing users into the dynamic of total expression. ”

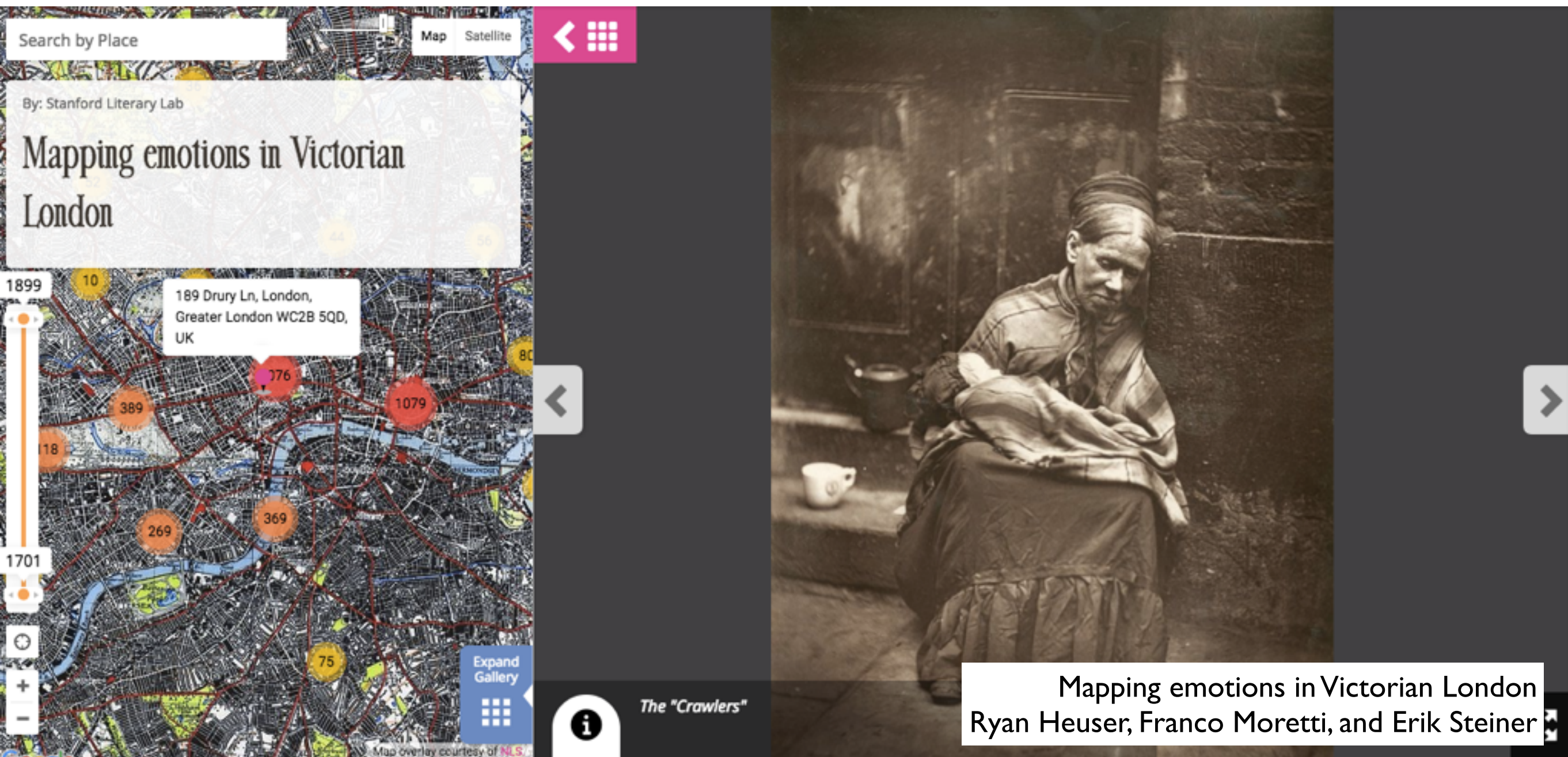
Applications for Academic Work



Diffusion of Lexical Change in Social Media
Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, and Eric P. Xing

“Rather than moving towards a single unified “netspeak” dialect, language evolution in computer-mediated communication reproduces existing fault lines in spoken American English.”

Applications for Academic Work



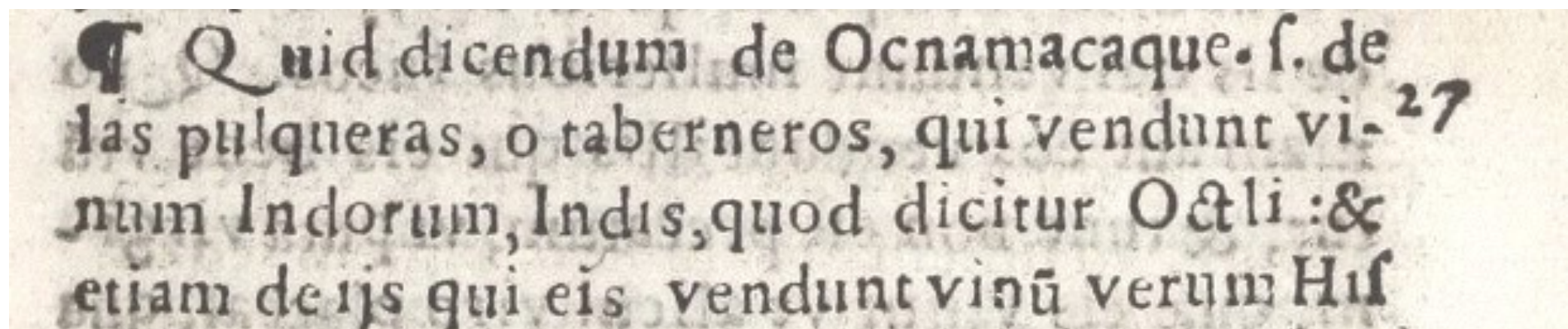
“We discovered a striking discrepancy between real and fictional geography, while also sketching the first lineaments for a future “semantics of space.””

| | | | | |
|---|--|--|--|--|
| <p>¶ Nic. lozque ininemitico in tototecuyo Je u x\~po: int quimochihuilico in quimotemachtili inictehuantin tictotepotztoquilizque nitemachtiltzin. ¶ No. Catlehuatl in quicuiloque in nahuintin in Euange tame.</p> <p>¶ Nic. Ca yehuatl yn auhtlamantli yn Euangelio yn ipan mopia yn ixq totech monequi in ticneltocazque y chihuaz\~q. Yhu\~a intleyn tictlalca íntictelchihuazque. Yhuan intley tic huaz\~q yye yx\'qch totech moneq yntlacamo ticneltocazque in tlein q cuilotiaque çan nim\~a ahueltitoma tizqui.</p> <p>¶ No. nicmatiznequi yn ic tiazque y huicac cuix ça yeyyo ticneltocazqu oquicuilotiaque.</p> <p>¶ Nic. ca \~itlacamo tictopielic\~a t milztican initeotenahuatiltzin, yyeh tin inquicahtiaque caniman ahuel que yn ilhuicac.</p> | <p>lozque ininemitico in tototecuyo Je u x\~po: int quimochihuilico in quimotemachtili inictehuantin tictotepotztoquilizque nitemachtiltzin. ¶ No. Catlehuatl in quicuiloque in nahuintin in Euange tame.</p> <p>¶ Nic. Ca yehuatl yn auhtlamantli yn Euangelio yn ipan mopia yn ixq totech monequi in ticneltocazque y chihuaz\~q. Yhu\~a intleyn tictlalca íntictelchihuazque. Yhuan intley tic huaz\~q yye yx\'qch totech moneq yntlacamo ticneltocazque in tlein q cuilotiaque çan nim\~a ahueltitoma tizqui.</p> <p>¶ No. nicmatiznequi yn ic tiazque y huicac cuix ça yeyyo ticneltocazqu oquicuilotiaque.</p> <p>¶ Nic. ca \~itlacamo tictopielic\~a t milztican initeotenahuatiltzin, yyeh tin inquicahtiaque caniman ahuel que yn ilhuicac.</p> | <p>lozque ininemitico in tototecuyo Je u x\~po: int quimochihuilico in quimotemachtili inictehuantin tictotepotztoquilizque nitemachtiltzin. ¶ No. Catlehuatl in quicuiloque in nahuintin in Euange tame.</p> <p>¶ Nic. Ca yehuatl yn auhtlamantli yn Euangelio yn ipan mopia yn ixq totech monequi in ticneltocazque y chihuaz\~q. Yhu\~a intleyn tictlalca íntictelchihuazque. Yhuan intley tic huaz\~q yye yx\'qch totech moneq yntlacamo ticneltocazque in tlein q cuilotiaque çan nim\~a ahueltitoma tizqui.</p> <p>¶ No. nicmatiznequi yn ic tiazque y huicac cuix ça yeyyo ticneltocazqu oquicuilotiaque.</p> <p>¶ Nic. ca \~itlacamo tictopielic\~a t milztican initeotenahuatiltzin, yyeh tin inquicahtiaque caniman ahuel que yn ilhuicac.</p> | <p>lozque ininemitico in tototecuyo Je u x\~po: int quimochihuilico in quimotemachtili inictehuantin tictotepotztoquilizque nitemachtiltzin. ¶ No. Catlehuatl in quicuiloque in nahuintin in Euange tame.</p> <p>¶ Nic. Ca yehuatl yn auhtlamantli yn Euangelio yn ipan mopia yn ixq totech monequi in ticneltocazque y chihuaz\~q. Yhu\~a intleyn tictlalca íntictelchihuazque. Yhuan intley tic huaz\~q yye yx\'qch totech moneq yntlacamo ticneltocazque in tlein q cuilotiaque çan nim\~a ahueltitoma tizqui.</p> <p>¶ No. nicmatiznequi yn ic tiazque y huicac cuix ça yeyyo ticneltocazqu oquicuilotiaque.</p> <p>¶ Nic. ca \~itlacamo tictopielic\~a t milztican initeotenahuatiltzin, yyeh tin inquicahtiaque caniman ahuel que yn ilhuicac.</p> | <p>lozque ininemitico in tototecuyo Je u x\~po: int quimochihuilico in quimotemachtili inictehuantin tictotepotztoquilizque nitemachtiltzin. ¶ No. Catlehuatl in quicuiloque in nahuintin in Euange tame.</p> <p>¶ Nic. Ca yehuatl yn auhtlamantli yn Euangelio yn ipan mopia yn ixq totech monequi in ticneltocazque y chihuaz\~q. Yhu\~a intleyn tictlalca íntictelchihuazque. Yhuan intley tic huaz\~q yye yx\'qch totech moneq yntlacamo ticneltocazque in tlein q cuilotiaque çan nim\~a ahueltitoma tizqui.</p> <p>¶ No. nicmatiznequi yn ic tiazque y huicac cuix ça yeyyo ticneltocazqu oquicuilotiaque.</p> <p>¶ Nic. ca \~itlacamo tictopielic\~a t milztican initeotenahuatiltzin, yyeh tin inquicahtiaque caniman ahuel que yn ilhuicac.</p> |
|---|--|--|--|--|

sites.utexas.edu/firstbooks

Applications for Academic Work

"¶ Quid dicendum de Ocnamacaque. f. de
las pulqueras, o taberneros, qui vendunt vi-
num indorum, indis, quod dicitur Octli .&
etiam de iis qui eis venduntvr nō verum Hif



Replication is sensitive to historical context.

Applications for Academic Work

recebida a nova com as mesmas demonstrações, & alvoroço

Diplomatic.: recebida a nova com as mesmas demonstrações, & alvoroço

Normalized: recebida a nova com as mesmas demonstrações, & alvoroço

Replication is sensitive to modern expectations.

Applications for Academic Work

*“...la verdad, cuya madre es la **hif**tor**ia** emula del **ti**ẽpo, **de**pofito de las acciones, **te**ftigo de lo **pa**ffado, **ex**emplo, y **au**ĩso de lo **pre**fente, **adu**ertencia de lo por venir.”*

-- Miguel de Cervantes
1605 Imprint (Quijote interactivo)

*“... la verdad, cuya madre es la **h**istor**ia**, émula del **ti**empo, **de**pó**s**ito de las acciones, **te**stigo de lo **pa**sado, **ej**emplo y **av**iso de lo **pre**sente, **ad**vertencia de lo por venir.”*

-- Jorge Luis Borges
Pierre Menard, Autor del Quijote

Texts transform as they replicate.

Digital textual studies allows us to:

Think critically about the digital age.

Use digitization to think critically about language and text.