



Les humanitats digitals a l'aula de català

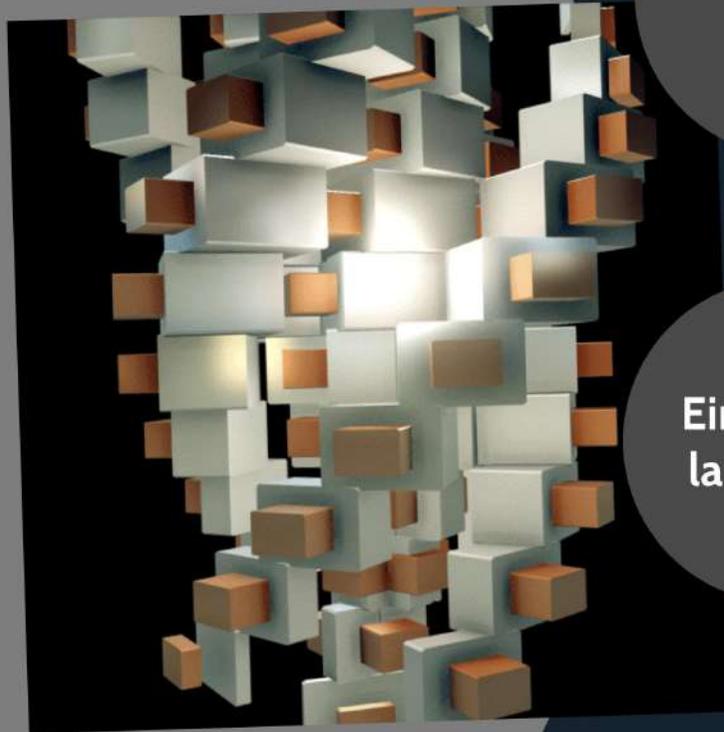
Dr Pedro Nilsson-Fernàndez
University College Cork, Éire

Som-hi!

<https://shorturl.at/elxDX>

Estructura del taller

1. Qui soc jo i què són les humanitats digitals?
2. Eines d'HHDD i la llengua
3. Propostes per a l'aula de català



Qui?

Què?

Eines HD i
la llengua

Propostes



Coláiste na hOllscoile Corcaigh
University College Cork, Ireland

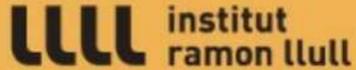
Digital Arts & Humanities
Ealaíona agus Daonnachtaí
Digiteacha

@UCC_DH

Recerca



University College Cork, Ireland
Coláiste na hOllscoile Corcaigh



Catalan Language and Culture

@CatalanUCC



University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Department of
Spanish, Portuguese
& Latin American Studies

@UCCSPLAS

1980

Estructura del idioma

1. Del latín, por el 1948, 400 de los hablantes "glotón"
2. Pícnico de 18-000 de los hablantes
3. Proposición por el 1948 de un idioma de cultura

Mapejant la Catalunya de Pedrolo

Recerca postdoctoral

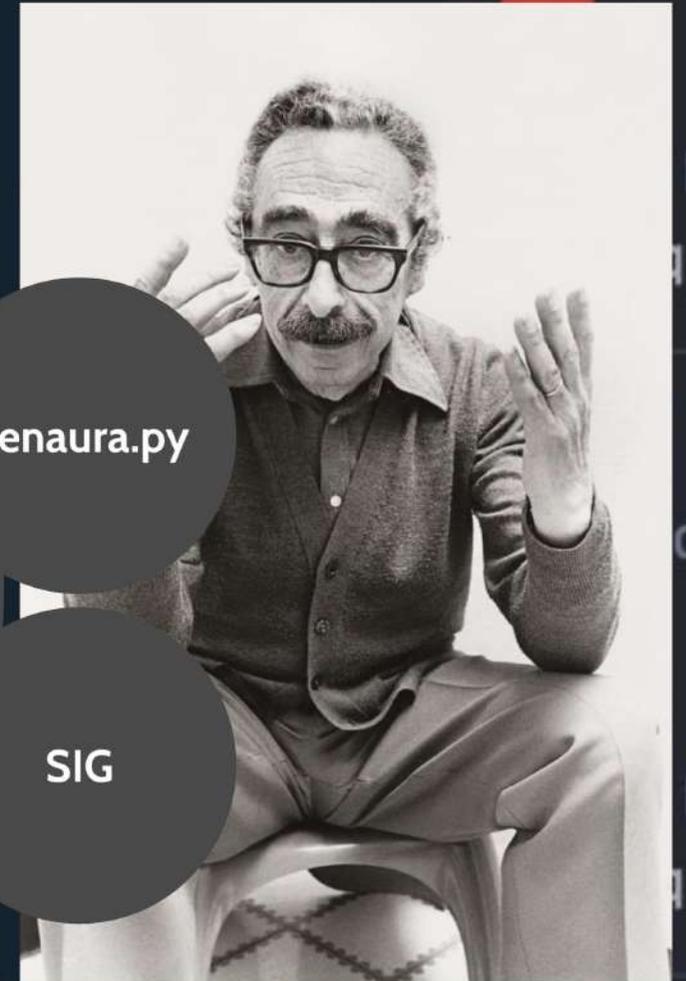
El multilingüisme a les humanitats digitals

Benaura.py

SIG

Multilingual
DH

Manuel de Pedrolo
(1918-1990)



Mapejant la Catalunya de Pedrolo

Recerca tradicional: historiografia, crítica literària, estudis culturals, estudis de gènere, teoria postcolonial.

Recerca postdoctoral

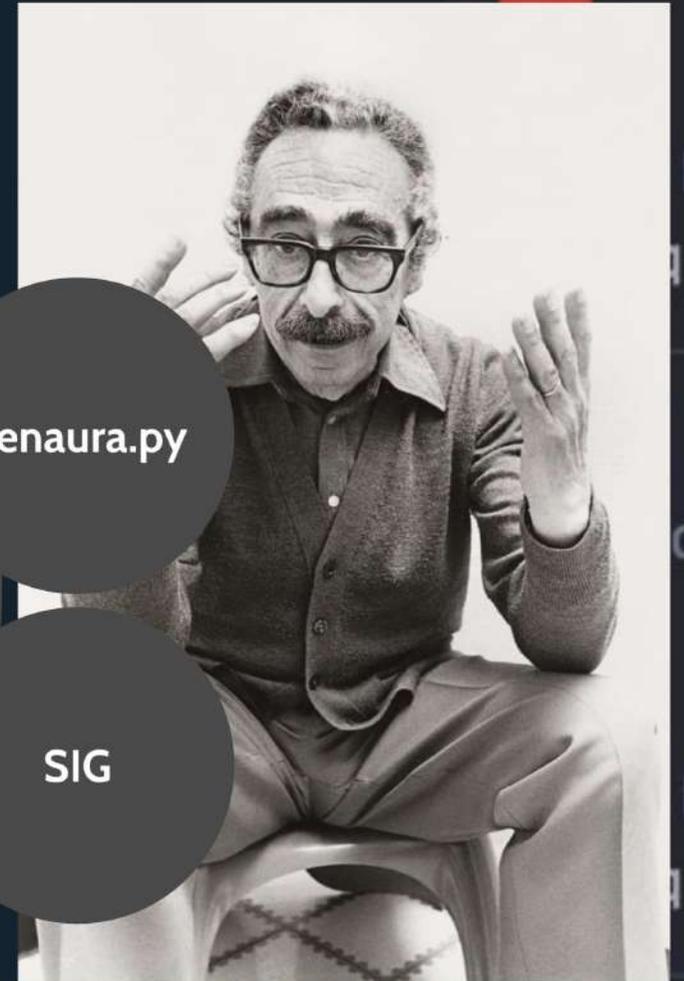
El multilingüisme a les humanitats digitals

Benaura.py

SIG

Multilingual
DH

Manuel de Pedrolo
(1918-1990)



Mapejant la Catalunya de Pedroló

Recerca tradicional: historiografia, crítica literària, estudis culturals, estudis de gènere, teoria postcolonial.

Recerca en HD: mineria de textos, lectura distant, creació d'un script dedicat al corpus de Pedroló (Benaura.py), visualització de dades amb SIG.

Recerca postdoctoral

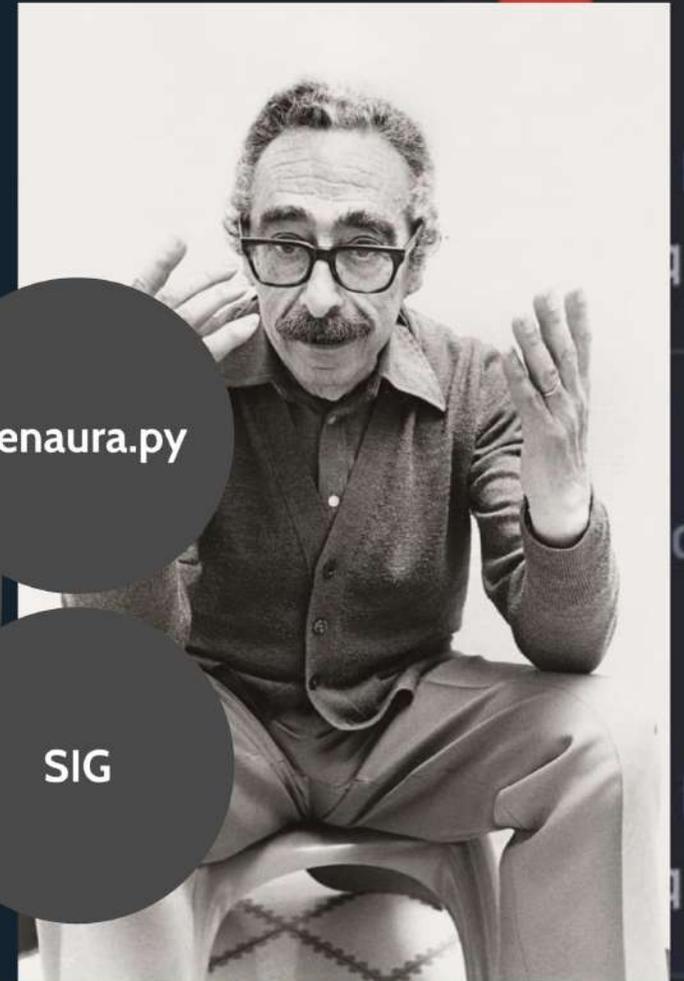
El multilingüisme a les humanitats digitals

Benaura.py

SIG

Multilingual
DH

Manuel de Pedroló
(1918-1990)



Mapejant la Catalunya de Pedroló

Combinació de:

Recerca tradicional: historiografia, crítica literària, estudis culturals, estudis de gènere, teoria postcolonial.

Recerca en HD: mineria de textos, lectura distant, creació d'un script dedicat al corpus de Pedroló (Benaura.py), visualització de dades amb SIG.

Recerca postdoctoral

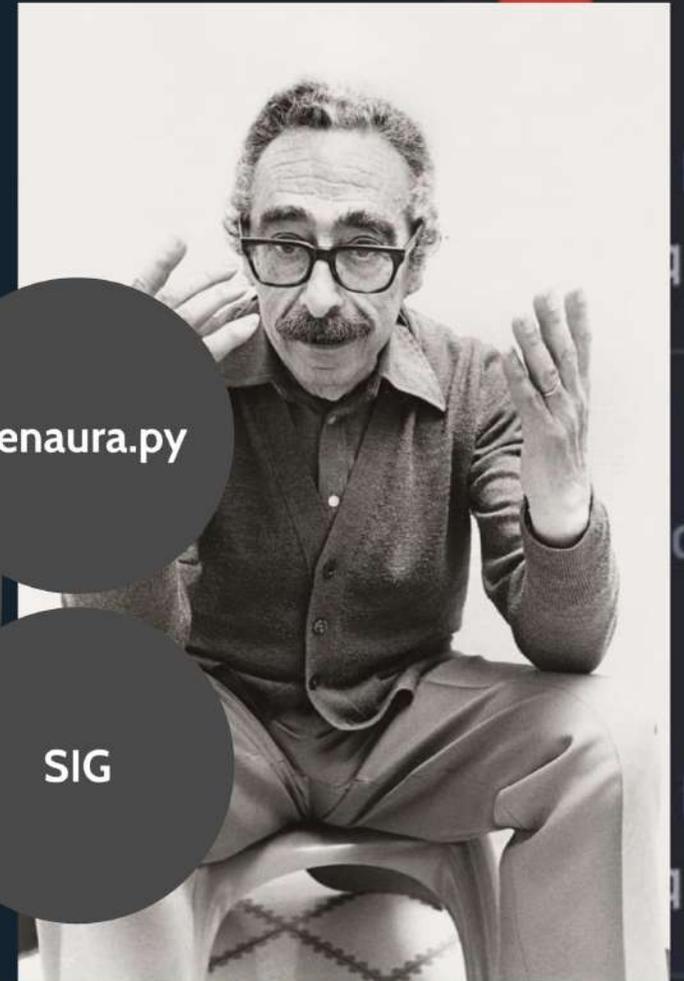
El multilingüisme a les humanitats digitals

Benaura.py

SIG

Multilingual
DH

Manuel de Pedroló
(1918-1990)



Benaura.py

Un script creat en python per al minatge de topònims

However, the tool proved to be more useful for an English-language corpus rather than a Catalan one, as Catalan words that happened to be homonyms of English language toponyms were identified by TopoText as locations around the globe. For these reasons, and in order to produce a more controlled environment in which to draw a more accurate picture of Pedroló's corpus in relation to the landscapes involved in this thesis, it became necessary to devise a unique script specifically designed for this project. In close collaboration with Dr James O'Sullivan – digital humanist and lecturer at the Digital Arts and Humanities Department in University College Cork – Benaura.py was created, a script allowing the extraction from a given Catalan text of toponyms referring to names of streets, squares, areas, neighbourhoods, cities, towns, etc., which had been previously compiled in a customisable database. The creation of this database involved a different set of operations; from a shapefile²¹ of the Iberian Peninsula downloaded from openstreetmap²² data providers Geofabrik, I was able to select the data corresponding to the Catalan territory using

- 20 Geoparsing is the process of converting free-text descriptions of places (such as 'Springfield') into unambiguous geographic identifiers (such as lat-lon coordinates). A geoparser is a tool that helps in this process.
- 21 For a full definition and history of the shapefile format, please refer to the article *Understanding Topology and Shapefiles* by David M. Theobald, Colorado State University.
- 22 OpenStreetMap is an initiative to create and provide free geographic data, such as street maps, to anyone. The OpenStreetMap Foundation is an international not-for-profit organization supporting, but not controlling, the OpenStreetMap Project. It is dedicated to encouraging the growth, development and distribution of free geospatial data and to providing geospatial data for anyone to use and share (OpenStreetMap Foundation website).

40

	A	B	C	D	E	F
1	217487284	A. Vives	A. Vives	BCN	41.387635	2.099309
2	32023920	Abadessa Olzet	Abadessa Olzet	BCN	41.392594	2.111794
3	22416258	Abaixadors	Abaixadors	BCN	41.383009	2.181049
4	201282163	Abat Escarri	Abat Escarri	BCN	41.449825	2.237579
5	46837964	Abat Odó	Abat Odó	BCN	41.432185	2.189314
6	239838074	Abat Safont	Abat Safont	BCN	41.375017	2.169809
7	35768197	Abat Samsó	Abat Samsó	BCN	41.396889	2.125093
8	123165190	Abat Teodor	Abat Teodor	BCN	41.429046	2.103087
9	407777029	Abd el-Kader	Abd el-Kader	BCN	41.41366	2.163912
10	115325240	Abdó Terradas	Abdó Terradas	BCN	41.400378	2.163557
11	82830767	Abedul	Abedul	BCN	41.366196	2.110017
12	3152496308	Aberra	Aberra	CAT	41.516819	1.9015635
13	90063294	Acícies	Acícies	BCN	41.423138	2.178243
14	290661144	Accés a la Finca Agustí Pedro i Pons	Accés a la Finca Agustí Pedro i Pons	BCN	41.406037	2.114821
15	545208361	Accés als Jardins de Jaume Perich	Accés als Jardins de Jaume Perich	BCN	41.393633	2.173481
16	12326975	Accés Llotja	Accés Llotja	BCN	41.378661	2.098876
17	36294748	Açores	Açores	BCN	41.425156	2.189849
18	385049144	actor Morano	actor Morano	BCN	41.413631	2.105585
19	290882628	actriu Tubau	actriu Tubau	BCN	41.413781	2.105192
20	5111090	Adolf Florca	Adolf Florca	BCN	41.384736	2.114161

Figure 3. Detail of database placename.csv

	A	B	C	D	E	F	G	H
1	city	lat	lng	freq				
2	anacapri	40.5558	14.2138	1				
3	Badalona	41.444611	2.2379545	1				
4	badalona	41.4494	2.2482	1				
5	Barcelona	41.384741	2.1829768	12				
6	Barra de Ferr	41.384619	2.179916	1				
7	Benaura	41.647	1.1412	12				
8	biblioteca de	41.38102	2.16952	1				
9	Borrell	41.404595	2.205554	1				
10	Gran Café Ca	41.37745	2.17674	1				
11	Cabres	41.382053	2.171444	1				
12	Cabrils	41.525728	2.3686084	1				
13	Camp d'aviac	41.3015	2.0801	1				
14	Camp de la B	41.416378	2.225946	1				
15	Can Tunis	41.3489	2.1547	1				
16	Capri	40.55	14.2246	1				
17	Carrer Ample	41.380057	2.179674	1				
18	Carrer Major	41.45133	2.202885	1				
19	castell de l'oi	40.82787	14.24775	1				
20	Castelldefels	41.265359	1.9543445	1				

Figure 4. Detail of output database mecanoscrit.csv

Mapejant la Catalunya de Pedrolo

Combinació de:

Recerca tradicional: historiografia, crítica literària, estudis culturals, estudis de gènere, teoria postcolonial.

Recerca en HD: mineria de textos, lectura distant, creació d'un script dedicat al corpus de Pedrolo (Benaurea.py), visualització de dades amb SIG.

Recerca postdoctoral

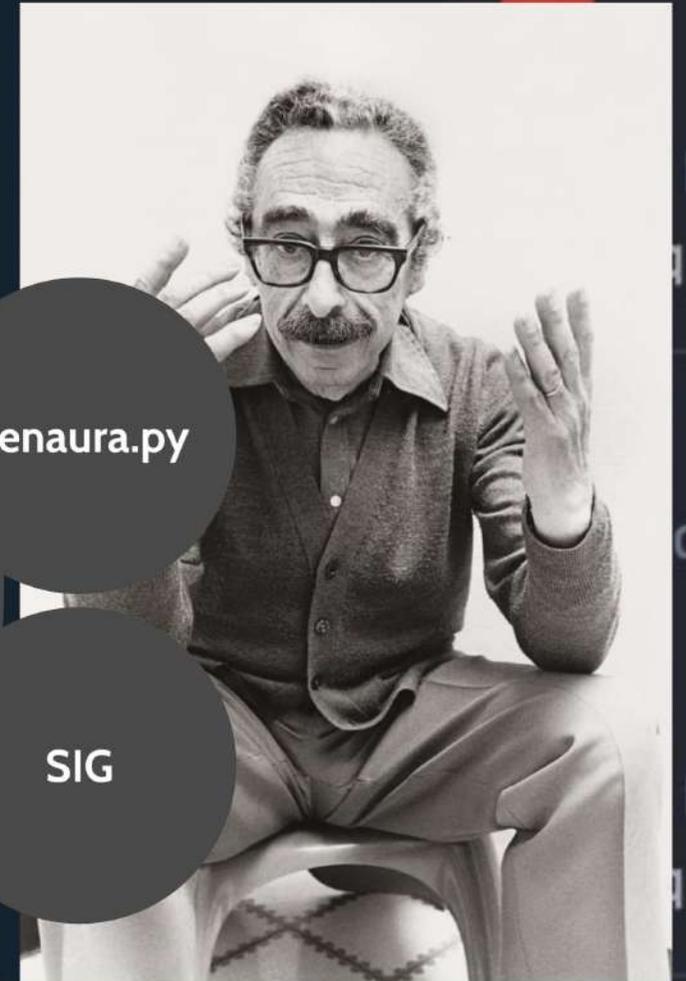
El multilingüisme a les humanitats digitals

Benaurea.py

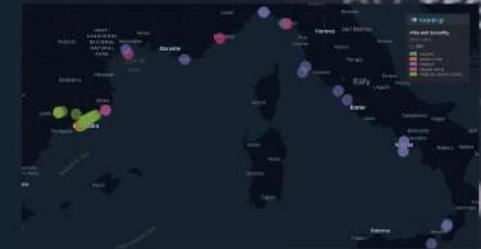
SIG

Multilingual
DH

Manuel de Pedrolo
(1918-1990)



Visualizació de novel·les de Manuel de Pedrolo amb Kepler.gl



MULTILINGUAL DH

Quinn Dombrowski @quinnanya
Academic Technology Specialist
Division of Literatures, Cultures, and Languages
Stanford University



Radius F
by freq

1970

Point co

Radius F
by freq

1980

Point co

MULTILINGUAL DH

Quinn Dombrowski @quinnanya
Academic Technology Specialist
Division of Literatures, Cultures, and Languages
Stanford University



CHAPTER EIGHT

Multilingual Digital Humanities

PEDRO NILSSON-FERNÁNDEZ (UNIVERSITY COLLEGE CORK) AND
QUINN DOMBROWSKI (STANFORD UNIVERSITY)

Digital humanities is contextual, shaped by the identities and interests of the people who engage with it, and the communities in which they are situated. The singular origin story featuring a meeting of minds between Father Busa and IBM has been complicated (Terras and Nyhan 2016) and called into question (Earhart 2015) and challenged by narratives that illustrate the diverse paths along which digital humanities has evolved (e.g., O'Sullivan 2020; Wang et al. 2020). The reasons motivating any one particular course of development vary: funding availability, cultural priorities, and the individual interests and resources available to influential early figures all play a role in shaping how a particular kind of DH came to be. Articulating the long-standing nature and diverse trajectories of DH communities beyond the mono-myth of Busa is an important ongoing project in the field, but even the Busa-linked lineage of "institutional DH" has its under-explored facets.

Radius F
by freq

1970
Point co

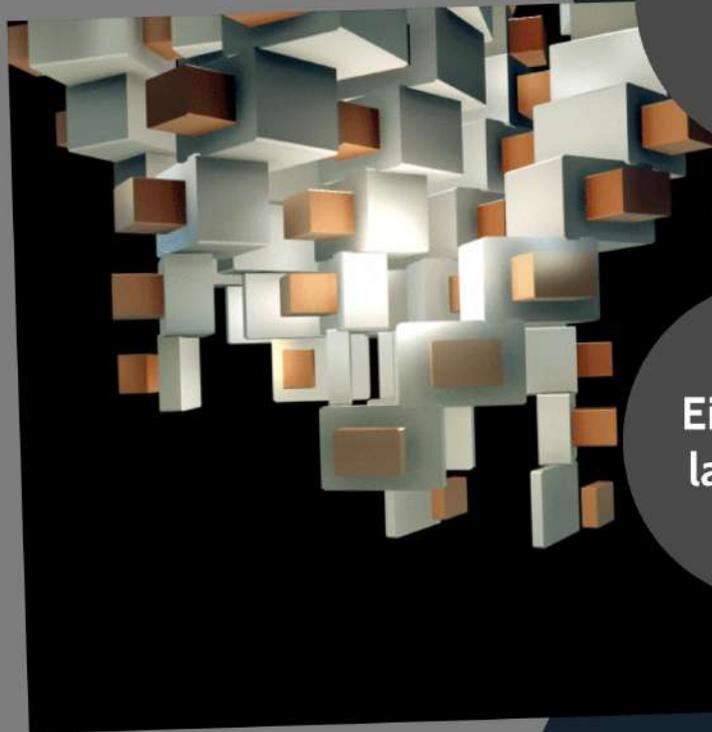
Radius F
by freq

1980

A small thumbnail version of the current slide, showing the title "MULTILINGUAL DH" and the authors' names.

Estructura del taller

1. Qui soc jo i què són les humanitats digitals?
2. Eines d'HHDD i la llengua
3. Propostes per a l'aula de català



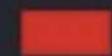
Qui?

Què?

Eines HD i
la llengua

Propostes

Què són les humanitats digitals?



Radius Rang
by freq

1970

Point color



Radius Rang
by freq

Sub-camps
de Les
HHDD

color

Què són les humanitats digitals?

Directora del centre d'Humanitats
Digitals a la Universitat de Michigan



“ *Digital humanities is a nexus of fields within which scholars use computing technologies to investigate the kinds of questions that are traditional to the humanities, or ... ask traditional kinds of humanities-oriented questions about computing technologies.* ”

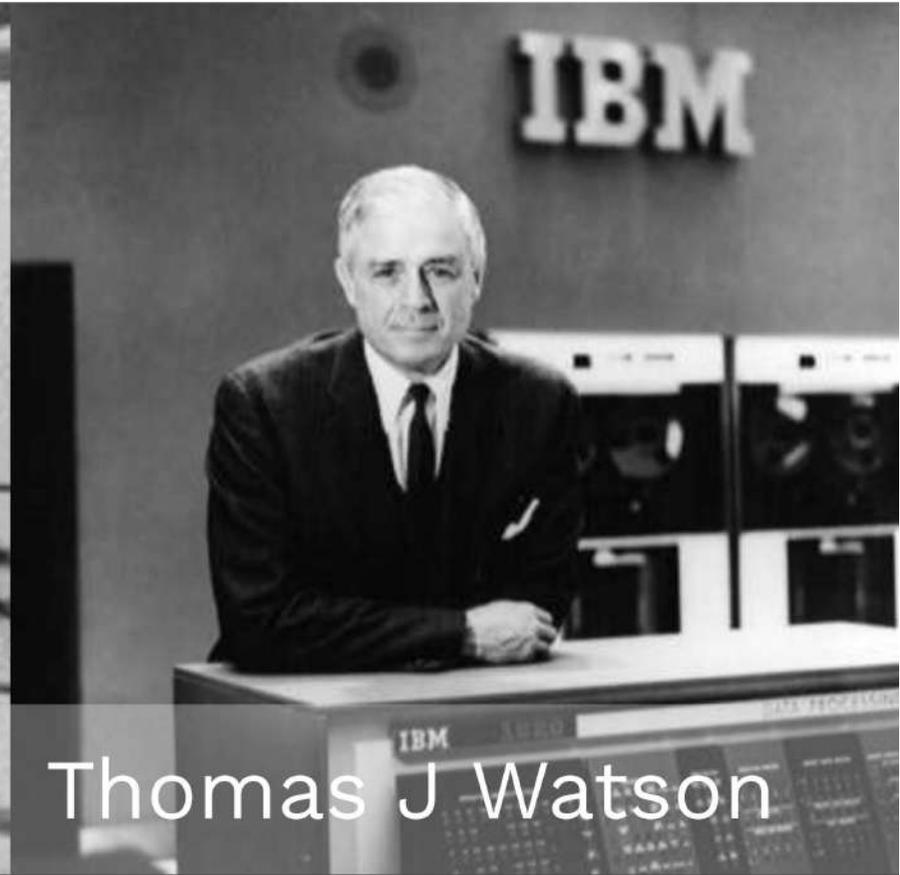
— Kathleen Fitzpatrick

Sub-camps
de Les
HHDS





Father Roberto Busa



Thomas J Watson



DHSI

DIGITAL HUMANITIES SUMMER INSTITUTE

Radius Rang
by freq

color

s Rang

by freq

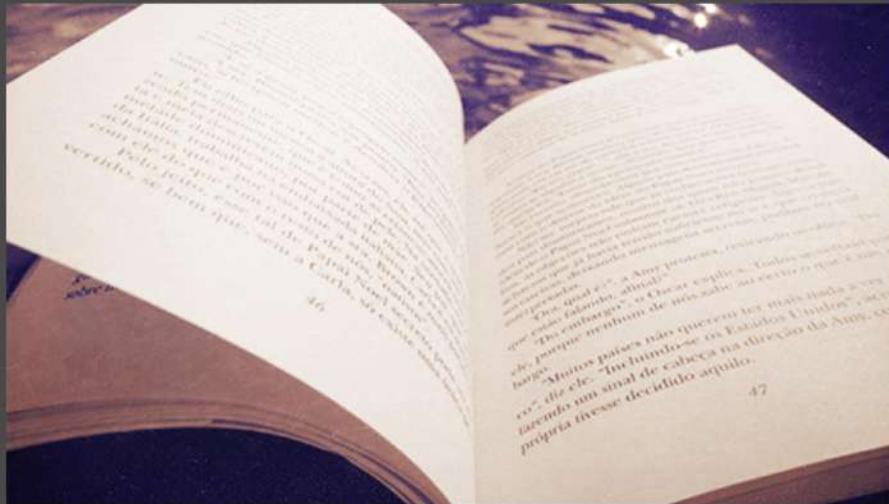
1980



Són les HHDD una sub-disciplina de les Humanitats?

- Audio/visual data processing
- Data Visualisation
- TEI (Texting Encoding Initiative)
- GIS (Spatial Humanities)
- Social Media Mining
- Open Access
- Cultural Heritage
- Digital Archives
- Crowdsourcing
- Digital Pedagogy
- Sustainability
- Gamification
- Virtual Reality
- Computer-assisted Text Analysis
- NPL (Natural Language Processing)
- Linked Open Data
- Digital Editions
- Digitisation

Subdisciplines de les Humanitats Digitals on el text ocupa un lloc central



Text
Analysis

TEI

GIS

Radius R
by freq

1970

Point co



Radius R
by freq

1980

Point co



Si hi ha una tasca que els ordinadors poden fer més ràpid que els humans és llegir/processar textos. Una gran part de les HHD es dedica a estudiar textos utilitzant les anomenades "lectures distants", un terme emprat per primer cop per Franco Moretti (2000).

Topic Modelling

Estilometria

Radius
by freq

1970

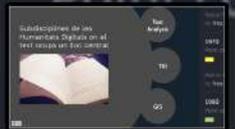
Point co



Radius
by freq

1980

Point co



Estilometria

L'anàlisi estilomètric de textos permet capturar l'estil d'una escriptora o escriptor, i per tant comparar-lo amb textos d'altres autores, o intentar atribuir autoria a obres anònimes.

Voyant
Tools

Radius
by freq

1970

Point co



Radius
by freq

1980

Point co



Estilometria

L'anàlisi estilomètric de textos permet capturar l'estil d'una escriptora o escriptor, i per tant comparar-lo amb textos d'altres autors, o intentar atribuir autoria a obres anònimes.



Volume 36, Issue 2
June 2021

JOURNAL ARTICLE

Who wrote *Wuthering Heights*?

[Get access >](#)

Rachel McCarthy, James O'Sullivan ✉

Digital Scholarship in the Humanities, Volume 36, Issue 2, June 2021,
Pages 383–391, <https://doi.org/10.1093/lhc/fqaa031>

Published: 26 June 2020 **Article history** ▼

“ Cite 🔑 Permissions 🔄 Share ▼

Voyant
Tools

Estilometria

L'anàlisi estilomètric de textos permet capturar l'estil d'una escriptora o escriptor, i per tant comparar-lo amb textos d'altres autors, o intentar atribuir autoria a obres anònimes.



Volume 36, Issue 2
June 2021

JOURNAL ARTICLE

Who wrote *Wuthering Heights*?

[Get access >](#)

Rachel McCarthy, James O'Sullivan ✉

Digital Scholarship in the Humanities, Volume 36, Issue 2, June 2021,
Pages 383–391, <https://doi.org/10.1093/lhc/fqaa031>

Published: 26 June 2020 **Article history** ▼

“ Cite Permissions Share ▼

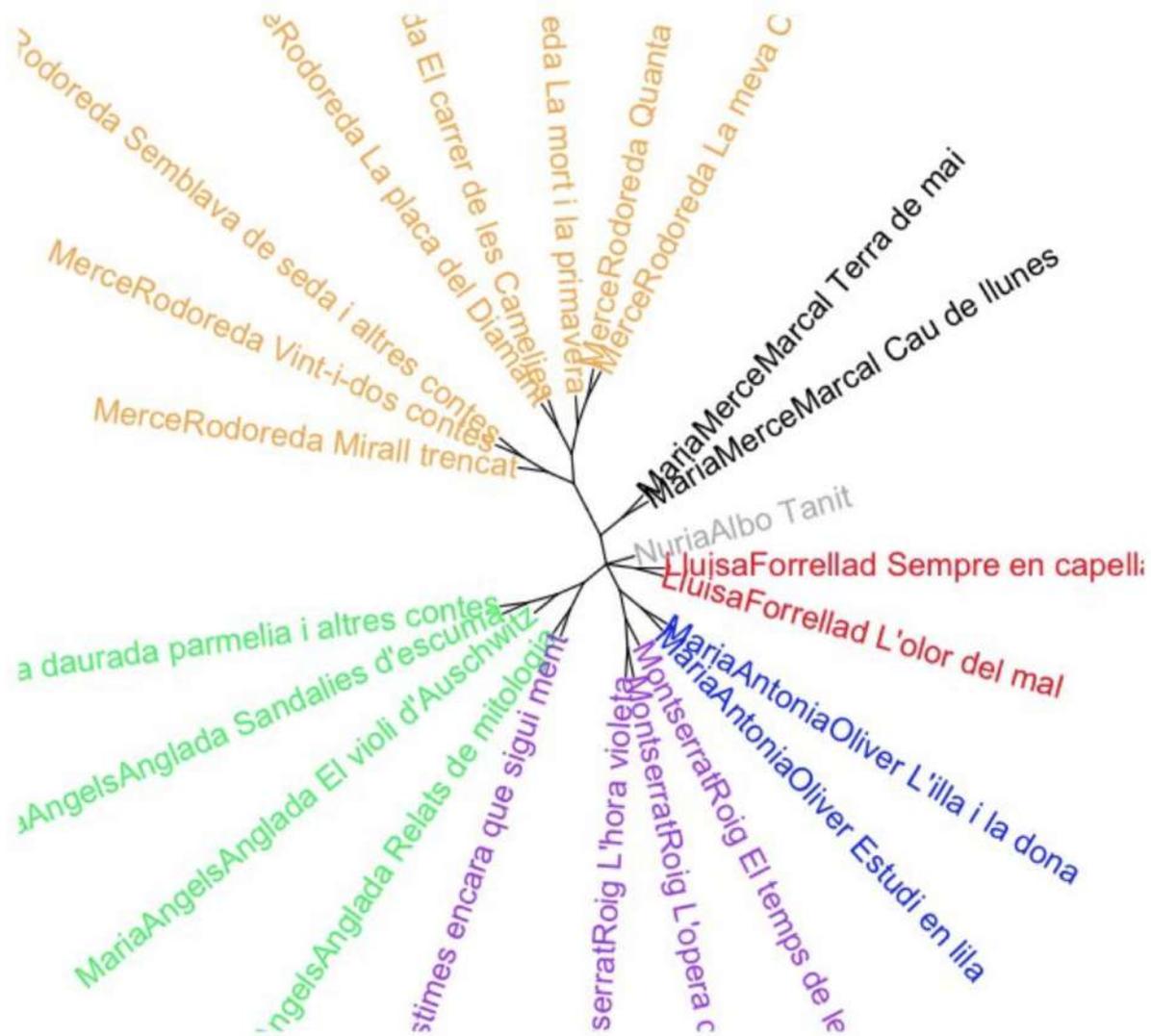
R Stylo package

Voyant
Tools

rar



Female Catalan Authors Bootstrap Consensus Tree



Radius
by freq

1970

Point co



Radius
by freq

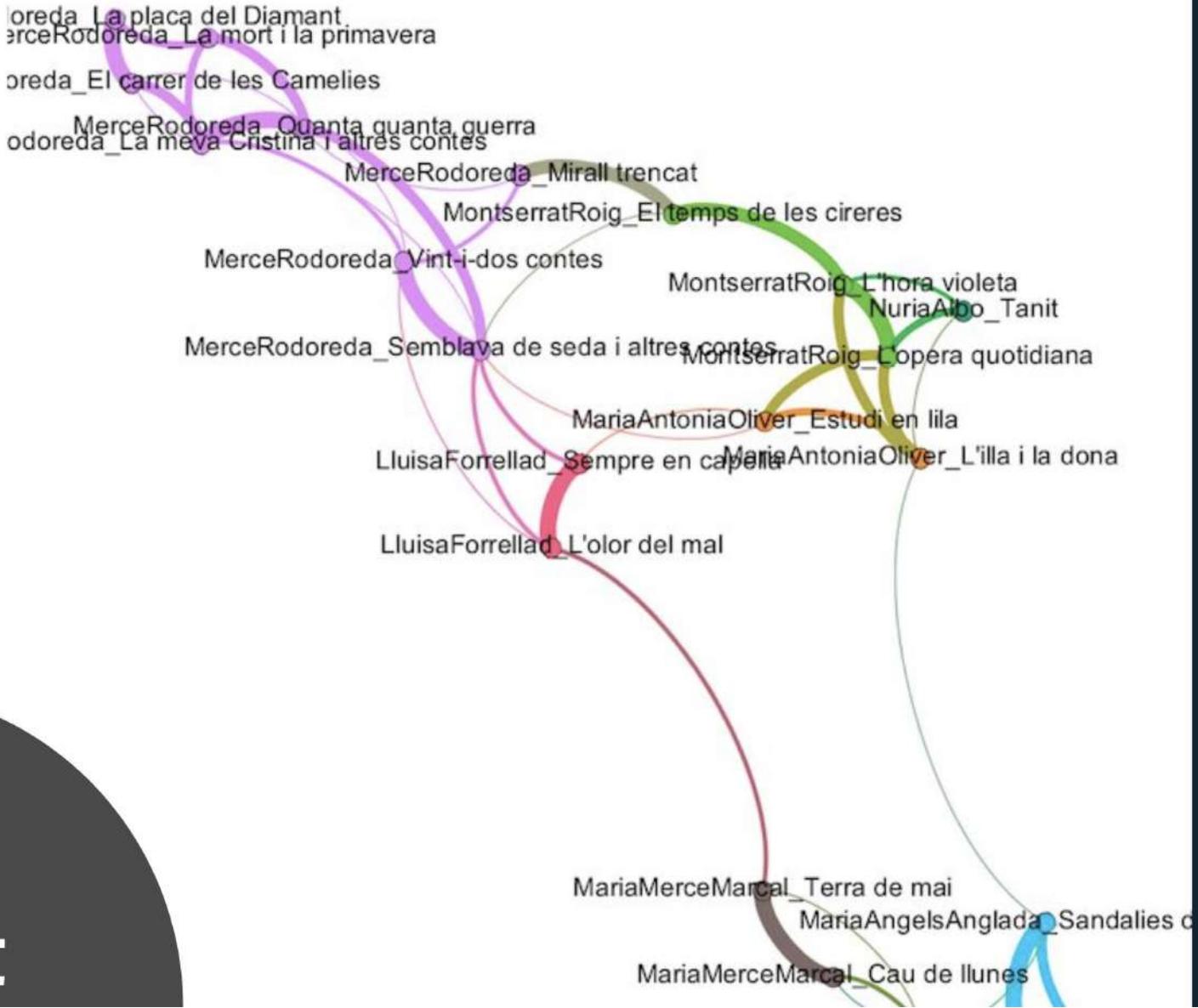
1980

Point co

Estilometria

Conte de la literatura de la nostra paraula catalana. És un gènere literari que s'ha desenvolupat a partir d'una tradició popular i que ha estat influenciat per la literatura clàssica i moderna.

4 slides package



SS >

e 2021,



Radius
by freq

1970
Point co

Radius
by freq

1980
Point co

Voyant Tools



Materials



<https://shorturl.at/elxDX>



Text encoding Initiative (TEI)

La Text Encoding Initiative (TEI) es un consorci que desenvolupa i te cura d'un estàndar per a la representació de textos en format digital.

LEAF
Writer

The image shows a side-by-side comparison of a text document and its corresponding TEI XML encoding. On the left, the rendered text is displayed with various elements highlighted in color: the recipient's name 'Patricia Spence' is in blue, the date 'October 21, 1935' is in orange, the location 'Stockholm' is in orange, and the time '21.10.35' is in pink. On the right, the XML code is shown, demonstrating how these visual elements are encoded using TEI tags. For example, the recipient's name is encoded as `<persName cert="low" ref="279399399">Patricia Spence</persName>`, the date as `<date cert="high" when="1935-10-21">21.10.35</date>`, and the location as `<placeName cert="high" ref="http://www.geonames.org/6453366">Oslo</placeName> to Bergen</p>`.

```
<telHeader>
  <fileDesc>
    <titleStm>
      <title>Sample Document Title</title>
    </titleStm>
    <publicationStm>
      <p/>
    </publicationStm>
    <sourceDesc sameAs="http://www.cwrc.ca">
      <p>Created from original research by members of CWRC/CSEC unless otherwise noted.</p>
    </sourceDesc>
  </fileDesc>
</telHeader>
<text>
  <div type="letter">
    <head>
      <title>Sample Letter - Bertrand Russell to <persName cert="low"
ref="279399399">Patricia Spence</persName> - October 21, 1935</title>
    </head>
    <opener>
      <note>
        <p>Bad writing due to shaky train</p><p>In train</p><p>
        <placeName cert="high" ref="http://www.geonames.org/6453366">Oslo</
placeName> to Bergen</p>
      </note>
      <dateline>
        <date cert="high" when="1935-10-21">21.10.35</date>
      </dateline>
      <salute>Dearest -</salute>
    </opener><p>I have had no letter from you since I left <placeName cert="high"
ref="http://www.geonames.org/2673722">Stockholm</placeName>, but I had a nice one
from John in an envelope you had sent him. I had sent him one addressed to
```

textos en format digital.

Sample Letter - Bertrand Russell to Patricia Spence - October 21, 1935

NOTE:

Bad writing due to shaky train

In train

Oslo to Bergen

21.10.35

Dearest -

I have had no letter from you since I left Stockholm, but I had a nice one from John in an envelope you had sent him. I had sent him one addressed to Copenhagen but he hadn't used it.

When I reached Oslo yesterday evening, Brynjulf Bull should have been there to meet me, but wasn't. He is not on the telephone, so I took a taxi to his address, which turned out to be a students' club with no one about on Sundays, so I went to a hotel feeling rather non-plussed. But presently he turned up. He had got the

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Sample Document Title</title>
    </titleStmt>
    <publicationStmt>
      <p/>
    </publicationStmt>
    <sourceDesc sameAs="http://www.cwrc.ca">
      <p>Created from original research by members of CWRC/CS C unless otherwise
    </sourceDesc>
  </fileDesc>
</teiHeader>
<text>
  <body>
    <div type="letter">
      <head>
        <title>Sample Letter - Bertrand Russell to <persName cert="low"
          ref="279399399">Patricia Spence</persName> - October 21, 1935</title>
      </head>
      <opener>
        <note>
          <p>Bad writing due to shaky train</p><p>In train</p><p>
            <placeName cert="high" ref="http://www.geonames.org/6453366">
              placeName> to Bergen</p>
        </note>
        <dateline>
          <date cert="high" when="1935-10-21">21.10.35</date>
        </dateline>
        <salute>Dearest -</salute>
      </opener><p>I have had no letter from you since I left <placeName cert="h
        ref="http://www.geonames.org/2673722">Stockholm</pl
        from John in an envelope you had sent him. I had se
```



<https://leaf-writer.leaf-vre.org/>

Radius f
by freq

1970

Point co



Radius f
by freq

1980

Point co



Text encoding Initiative (TEI)

La Text Encoding Initiative (TEI) es un consorci que desenvolupa i te cura d'un estàndar per a la representació de textos en format digital.

LEAF
Writer

The screenshot displays the LEAF Writer interface. On the left, a sample letter is shown with highlighted text and a date stamp. On the right, the corresponding TEI XML code is visible, showing the structure of the document and the encoding of the highlighted text.

Sample Letter - Bertrand Russell to Patricia Spence - October 21, 1935

NOTE:
Bad writing due to shaky train
In train
to Bergen

21.10.35

Dearest -

I have had no letter from you since I left Stockholm, but I had a nice one from John in an envelope you had sent him. I had sent him one addressed to Copenhagen but he hadn't used it.

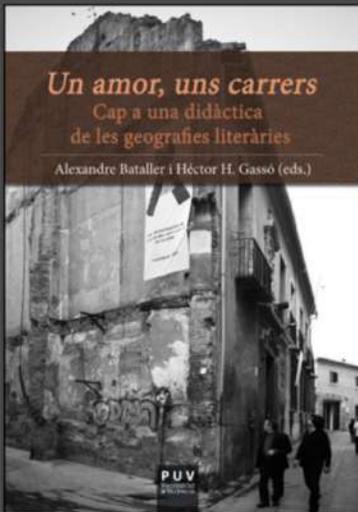
When I reached Oslo yesterday evening, Brynjulf Bull should have been there to meet me, but wasn't. He is not on the telephone, so I took a taxi to his address, which turned out to be a students' club with no one about on Sundays, so I went to a hotel feeling rather non-plussed. But presently he turned up. He had got the

```
<?xml version="1.0" encoding="UTF-8" ?>
<teiHeader>
  <fileDesc>
    <titleStm>
      <title>Sample Document Title</title>
    </titleStm>
    <publicationStm>
      <p/>
    </publicationStm>
    <sourceDesc sameAs="http://www.cwrc.ca">
      <p>Created from original research by members of CWRC/CSEC unless otherwise noted.</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
<text>
  <div type="letter">
    <head>
      <title>Sample Letter - Bertrand Russell to <persName cert="low" ref="http://www.geonames.org/279399399">Patricia Spence</persName> - October 21, 1935</title>
    </head>
    <opener>
      <note>
        <p>Bad writing due to shaky train</p><p>In train</p><p><placeName cert="high" ref="http://www.geonames.org/6453366">Oslo</placeName> to Bergen</p>
      </note>
      <dateline>
        <date cert="high" when="1935-10-21">21.10.35</date>
      </dateline>
      <salute>Dearest -</salute>
    </opener><p>I have had no letter from you since I left <placeName cert="high" ref="http://www.geonames.org/2673722">Stockholm</placeName>, but I had a nice one from John in an envelope you had sent him. I had sent him one addressed to
```



SIG (Sistemes d'Informació Geogràfica)

L'evolució digital de l'estudi de les geographies literàries, per exemple, passa per emprar els SIG per representar espais literaris.



Kepler.gl

Radius f
by freq

1970

Point co



Radius f
by freq

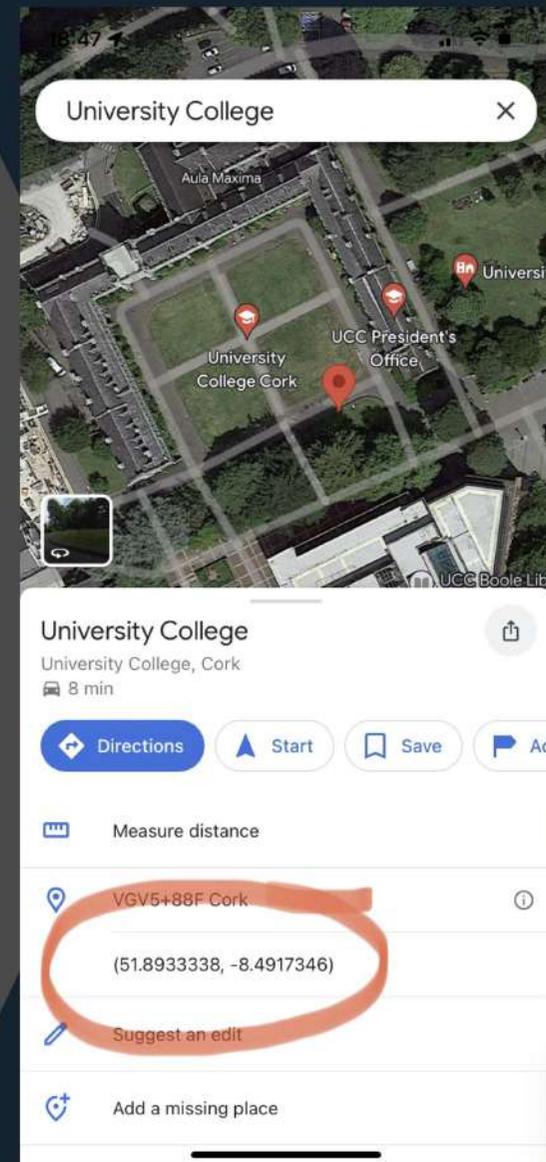
1980

Point co

On ensenyeu català?

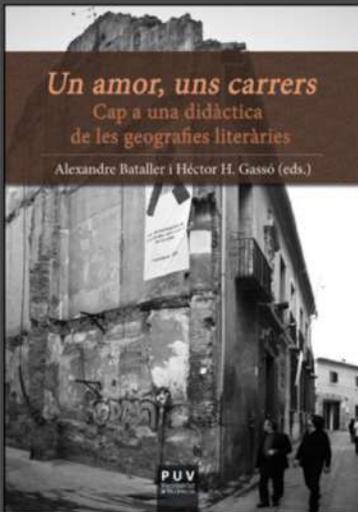


<https://forms.gle/hq5xW8Y57ZNzhFDYA>



SIG (Sistemes d'Informació Geogràfica)

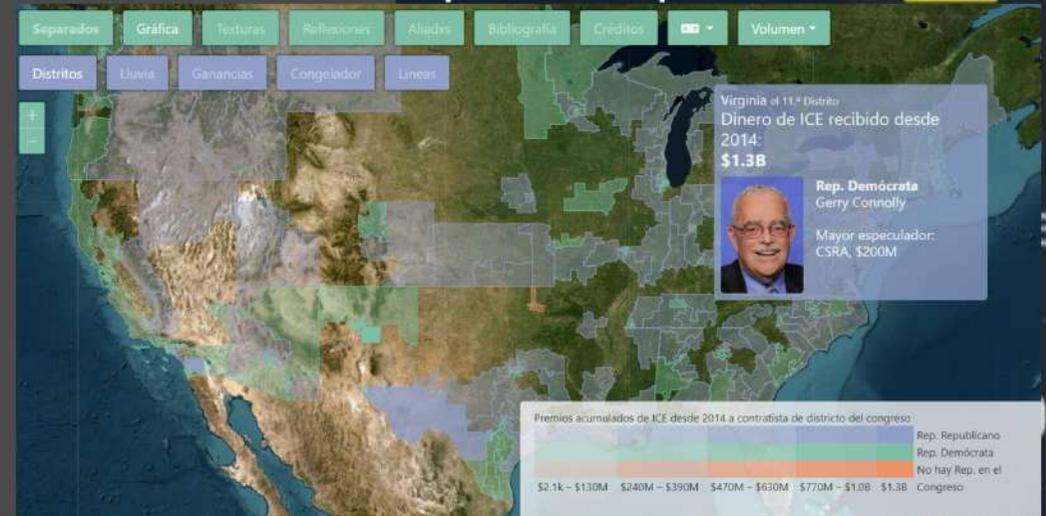
L'evolució digital de l'estudi de les geografies literàries, per exemple, passa per emprar els SIG per representar espais literaris.



Més enllà de la geografia com a camp d'estudi

Kepler.gl

Torn Apart/Separados



VOYANT TOOLS / TweetDeck

- Localitzar exemples de paraules, expressions o elements sintàctics que puguin generar dificultats a l'alumnat:
 - Deu-n'hi-do, malament rai,
 - Pronoms febles
 - Imperfecte vs perifràstic

TEI

Radius Range

freq

1970

Point color



Radius Range

by freq

1980

Point color



- Tots utilitzem unitats didàctiques que contenen textos amb un nombre considerable de referències històriques o biogràfiques, ja siguin noms d'esdeveniments, de personatges clau per a la cultura catalana, llocs, corrents estilístiques.

- La meua proposta es preparar un text curt en format xml que contingui aquestes referències que voleu que els alumnes s'enduguin a casa després de la lliçó.

- Amb LEAF, demaneu que els alumnes afegixin etiquetes als termes que considereu importants. LEAF converteix aquest text força pla en un hypertext, amb connexions amb wikipedia que els alumnes poden utilitzar per ampliar els seus coneixements sobre cada concepte.

Radius Rang
by freq

1970

Point color



Radius Rang
by freq

1980

Point color

