



Linking languages across space and time

John P. McCrae National University of Ireland Galway Insight Centre for Data Analytics



Biography





Postdoctoral Researcher 2009-2015

Universität Bielefeld



Lecturer 2015-



Over a 1000 datasets of which ~130 are of linguistic relevance

OBJ

Linguistic Linked

Open Data

(Linguistic) Linked Open Data

 Use URLs to identify things
Use URLs that tell users how to find your data

3. When someone looks up an URL, provide useful information

4. Include links to other datasets.



linghub.org

Central aggregator of metadata about language resources across repositories

sommittee

hub [hab] n a (of wheel) (Rad)n

Prêt-à-LLOD

Zentrum nt, Mittelpunkt m; a.

finanzielles Zentrum; the - of

hubble-bubble [h

Wasserseife f[b] (=

Radaum; a - of

n (idad') yodun

mouse Un-Amer

It; a ~ of activi

island's social

punkt der Ins

rt n (US) Gro

habab) n

Expanding to 1,000,000 resources

OntoLex

W3C Vocabulary for representing dictionaries as linked data.

https://www.w3.org/2016/05/ontolex/

Community group is still working, please join!

http://www.w3.org/community/ontolex









Global WordNet Association Formats

LexicalResource			
Lexicon+			
LexicalEntry+			
Lemma			
Form*		GWN-LMF	
Sense*			
Synset*	OntoLex	\times	JSON
Definition*			
SynsetRelation*			



Linking Dictionaries

 Activity involving mental or physical effort done in order to achieve a purpose or result. 'he was tired after a day's work in the fields'

+ More example sentences + Synonyms

Mental or physical activity as a means of earning income; employment. 1.1 'I'm still looking for work'

+ More example sentences + Synonyms

The place where one is employed. 1.2 'I was returning home from work on a packed subway

+ More example sentences

The period of time one spends in paid employment. 1.3 'he was going to the theatre after work'

+ More example sentences

1.4 West Indian [count noun] A job.

'I decided to get a work'

+ More example sentences + Synonyms

2 A task or tasks to be undertaken. 'they made sure the work was progressing smoothly'

+ More example sentences

+ Synonyms

- The materials for a task. 2.1 'she frequently took work home with her'

2.2 (in combination or with modifier works) British Activity involving construction or repair. 'extra costs caused by additional building works'

1. (heading, uncountable) Employment.

Labour, occupation, job. [quotations v]

My work involves a lot of travel.

2. The place where one is employed.

He hasn't come home yet, he's still at work.

3. One's employer

"I want to go to the R.E.M. reunion concert but I'm not sure if my work will let n

2. (heading, uncountable) Effort.

1. Effort expended on a particular task.

Holding a brick over your head is hard work. It takes a lot of work to write a di

Sustained human effort to overcome obstacles and achieve a result.

We know what we must do. Let's go to work.

3. Something on which effort is expended.

There's lots of work waiting for me at the office.

(physics) A measure of energy expended in moving an object; most commonly, for Work is done against friction to drag a bag along the ground.

5. (physics, more generally) A measure of energy that is usefully extracted from a pro-

3. Sustained effort to achieve a goal or result, especially overcoming obstacles. [quotations] We don't have much time. Let's get to work piling up those sandbags.

4. (heading) Product; the result of effort.

- 1. (uncountable, often in combination) The result of a particular manner of production There's a lot of guesswork involved.
- 2. (uncountable, often in combination) Something produced using the specified mate We've got some paperwork to do before we can get started. The piece was de
- 3. (countable) A literary, artistic, or intellectual production. [quotations v]
 - It is a work of art. the poetic works of Alexander Pope
- 4. (countable) A fortification.

William the Conqueror fortified many castles, throwing up new ramparts, bastic

- 5. (uncountable, slang, professional wrestling) The staging of events to appear as real.
- 6. (mining) Ore before it is dressed.

(Can we find and add a quotation of Raymond to this entry?)

⁺ More example sentences



Lexicography





S european lexicographic



Objectives:

- Support needs of lexicographic community
- Robust guidelines and best practices for workflows in lexicography
- New methods for retro-digitisation of lexicographic content



Interoperability and Linked (Open) Data

european lexicographic infrastructure



Objectives:

- Interoperability of existing lexical resources
- Establish cross-lingual links between dictionaries
- Validate the quality of dictionaries



Lexicography for NLP





Objectives:

- Show impact of quality lexicographic data on NLP
 - Word Sense Disambiguation
 - Entity Linking
 - Semantic Parsing
- Address "knowledge acquisition bottleneck" in AI



NLP for Lexicography

Objectives:

- Automatic acquisition of lexicographic data
- Crowdsourcing and gamification
- Enrich dictionaries with multi-modal data





elexis

european lexicographi infrastructure



How to link your dictionary





Bailiú Bagáiste agus Slí Amach Baggage Reclaim and Exit

Naisc Litilte

Step 1: Get two dictionaries and figure out which senses link



Step 2: Extract facts from the entries



Step 3: Textual Similarity

place for keeping money

An institution where one can place and borrow money and take care of financial affairs

Text Analysis/ Semantic Analysis/ Deep Learning



Step 3: Graph (contextual) Similarity



Step 4: Combine Features





All approaches improve F-Measure from 20% to about 60~75% (on English)

Resource-intensive and requires a lot of training

New benchmark and shared task on English, Dutch, Danish, Spanish, Italian, German, Slovene, Estonian, Bulgarian, Hungarian and Irish.

& your language? If you have **2 open monolingual** dictionaries and about 4-5 hours.



Minorit (ized) and Fate go Relative State go Room a Charge Cost 65.05 Ser Octve Cost 65.05 Ser Octve E 63.55 per Night Cost 65.05 Ser Octve E 63.55 per Night E Cosc ar Galway County Council **Oumpáil!** No Dumping Fineáil €1900.00 £1500 FINE € 1904.61



















Towards a dictionary of all languages





SACRANI

tendete obtantros ce. -- 2. a) mathaarrow in pressa crystel, Lieves in

sorous crewin, c) aneros, Satrant, drum re cospored, neur a Jaarryw. Sacrianus 3, companents, specialized, II of L. socto in Chapterante an company operation. - 2 conturned, span ; antrola Ditla monaci-

s

REAT CHEF. sarrate one carro, desconcerano. sacritur, dris m ocsetures. sacratus J. L. conurs templom. - 2. oferenzapes, oferance day,

sacrosu-sacri, non, pl, on sares, ascricula, as 'so solito oftenaca avertas, Marrie .

antrifer, fers, ferant accuts carness BORDERTS.

sacrificalla, e otracate co ana septate dosmochroners apparents.

ancriticatus J L epunctes a serpena-2. miles r Spontra weyras; sufat, sacrificati, trum re operatione wryt-AS (18 TOWETHING, NOWTH HOPALS, CYPER, specietate appres as connecks do-FORT &

secrificatio, dais / wepter-personancesacrificator, dris m andro provance # 22738.

sacrificium, II л нортял, жартанартautomete facere, tracffuere, sacrifices student

sarrilles I spession article, southynow succe, pecces.

sacrificalus, I et andro openance ment-RE. RIGHT PER SACHECRES WORK. andre alwapped at the wepterupped. TOTATA, ADATE IN-DATES PARADOMERAL

secrificus 3 L resources supres rex --2. STARTING CC. 20 MAPPARTONNAL CONTRACTOR

TO, Reprises dies, silva.

satelling& adv. festivates, rates, rates varrilegium, II or L a) carturaterau. by spin, consideration over

anerslagun 3 L. Cocrossitionen; ander acrilegus, 1 m cacturates, ficanes, aves at spanne. -- 2. bestown.

Conclusion

sacro I I. a) monetunan ma dora area. Isuram Phoebo. b) nocurroses, annaconcentrate, ofgeressi honorem allest altquem tells Exendri ; quod Libiting sarravit sorro e wapreo. - 2 a) comscanne, common as merginetemmen. as messapyones fordus; deum sede conveniences an five apare ; leges as gratue cecurity - Janmes, - where news HADDREND BOCCOU DERCH REALMOND & promitation, b) yeratemetation, ofercwagroups aliquem Lenhus plectes. menhoriza.

sacró-sinctus il menoranemiaca, cam agen, schopyterss postenale.

sacrafica-sperifica.

sacrum, I & L a) commun openant, car-TANK, STOTPORTONS NO DOC DECENTION. que tibi commendat Tanta penates. sacra en sedibus eripait, sacra Cereris, b) scents secret accendere -2 a) cathornellenetane, permission of para wrotanoowaccosmoe ascreen facere ; Giseco asens no rysume mipes b) spensies, raparetno; secre legule custiens represents tota gentiles gamman, cenedan wereacopumpernot. - 3. a) corner legations, regain DI TAUMTTRO, MONTOPHER IN TRANS. SHITE tori, Alteratory,

saeclum ele, sieculum.

saccularis, a croserce | indi mintu cu PERSONALISA SCENE CTO PORNAME ; CARDINE sacculate neces, amato neeza ana span-WINDOW TO DA CAND CUDATINE.

saccolom a sacclum, I = 1, nonplease white watched white (33 1-2 manual multe saccula hominum, aseclis elfeta senectus. - 2. a) ses, spawe, newsremot surra soccula, postri infamia, specult b) spewe, syx as specific otheran, sport mitescant specals. A DESCRIPTION OF THE OWNER

samps with (compt samples, superi, " anepissime) were; quant aveplatime REARDING MORE DO-90226. anepe-numero adu, secos, surple, is f 2, norr, orpass. 2. nors. sperpals. BREDIA -- SEDIA. sacpical & size, wectores. easpimentum, I # orpaga. Sacplaum, I a Cenesys, sparse a Cas-BHYS. sargio, saepsi, saepius 4 I, orponate c mort. -- 2. a) orpangane, anderaness undern muzis ; so bertis sensagess ce a asopena; spes, locum cogitatione of annual of mersons. By savanwas, saugenmans vice; arbers pricsidio norranse a rpana rapancos. c) operpawgas, ppeupesant onnes haas addition, transitions, uneptum, I a 1, orpana ; orpaneno warna ; moop enlist victima sarplis ; area. apercore Hanaversum war cany ampturn gaugeorys - 2 pl anyanous warro as Marcuso musi animu as Childrens 23 / LICENSEL gants (asta), so / 2, warms, morne faires proaders sorter; route son-CAR ADDAM - 2. POPER. 4) HOMES AS shows, b) wreak Sociable, is / Crisler, spon a vicepano. Sartabus. 3 cerufacion. sactiger, gera, gerum account and, pecus | Suddl. anetiger, art re rescan. Course ARADOWS - METURNET ADDL. ANTER ADD. CARDENS, MCCOMM. COURSE. allevidicus 3 manual rates, antidam-TRACKING. sarvio 4 convertagane, former, dull-CTRIERS IN ADQUARCE IN CONSUMER AN Tiberes. and willin, an o sarvities, di / L cuspemer. - 2. wertlauer, cypenser, coproce hostiam, indicis; annonar cal-INCOME. sarvitūdo, inis / mornance : espectersarvus 3 1, caspes Mo. -- 2, acrist. CYPOR. CYPANORA, YAACAM.

SAEPE

sign, se / 1. openerations, ranges, as, spoppings, -- Z. constructs.

sagăcităs, âtis / 1. a) mocrpeas diomanne, otrygener cartery, by synchronemory several ---- 2. sponsarraseer, screeywar howints.

sagicury our. L. r across of warme .---2. spen, speakantrate, toward,

Segulassus, I / Caratar, 1913 to Depanes, Segularedness 3 carameteran.

Sagaria, is a Songarius, It or Corspon. pena sue Dersson a Operas. Sagaritis, ids / carapatan,

regiltes à places a curs pres augures. sagis, get, -kets L c actue of constant (redworder) canis, - 2 spontante-AND, MERTING, OCTOVING SIGNA, ADgentia.

sagina, as f.L. room, prisants anneron. - 2, 43 abgut, speed, annufemrun, gladutaria, b) yours softthe candely segment, of provider, LINCTORS ADDRESS.

sagisates 3 organos, trues.

anging I process, expenses, apone porcum; anglest elepsets helps yrown.

slight A matterney, tornine.

REGILLA, BR / COMPANY PARAMETER COMPANY sagittarius, it manprova.

segitteres il contare cur cryone,

sagilitier, fera, ferum scorin crystas

Sagilli-potens, entils m Crysnes ics-

saging I crosses, available crowin.

segma, alle a fund cavary, trinsport

sagenina, am o caroless there (excas) vpeak astrongs ov Kauppingframe rais, touts demante southe chr. cole co a new as acognitic memory. sorgen ermsion apro opacchasics's

Sagra, as mf Curps, pris a Doymeye induscr a Rhuns Musical

aspalitus 3 officies a appulate.

sagalum, V.N. Thronestan Sprint and BRIDER DESC.

angum, I & L. cura, anfana, and speat. 2. hofwares apras, manna Apex somers supum, on od angu

Summary

- Dictionaries are a vital research tool for humanities and NLP
- Linking and integrating is challenging and important
- Link under-resourced languages as closely related languages
 - Historical \rightarrow Modern
 - Minority \rightarrow National





Thanks, questions?