

From print to digital -implications for dictionary policy and lexicographic conventions



20th International Conference
University of KwaZulu-Natal

Michael Rundell
Lexicography MasterClass and Macmillan Dictionaries

1. The dictionary landscape in 2015

- A turbulent phase - between the old and the new
- “The future is already here — it's just not very evenly distributed” (sci-fi writer William Gibson)
 - in some areas, digital media already dominant; in others, not yet widespread
- “Lexicographers ... currently live in a sort of interregnum” (Hanks 2015.87)
 - working in digital media is still (relatively) unfamiliar: what are the new rules?

The dictionary landscape in 2015

- An uneven picture
 - In some sectors, print dictionaries in rapid decline
 - e.g. English dictionaries for mother-tongue speakers, English learner's dictionaries, two-way bilinguals for “big” languages (e.g. English-Spanish)
 - In others, demand for printed dictionaries likely to persist
 - e.g. schools dictionaries, “smaller” languages, specialised dictionaries

The long-term trend

- Why dictionaries aren't like other books
 - “People typically consult maps, encyclopedias, and dictionaries *while they are doing something else*” (Nesi, forthcoming)
- For reference resources, digital is the ideal medium
 - no space limitations, continually updateable, multimedia opportunities, easily linked to other data sources, etc.

2. Traditional dictionary publishing

- Lexicography evolved in the age of print
 - for English, a 400-year-old model
- Goal: maximum information in finite space
 - broadest possible coverage of the lexicon
 - fullest possible information about word behaviour
 - a “zero-sum game”: if you add something new, you have to take something else out
- Much of what we take for granted as “normal” in dictionaries is driven by this imperative

The imperative to use space optimally

- Philip Gove and Merriam-Webster's *Third International*
- “Every editorial decision Gove made was dictated by space: the need to create as much of it as possible so he could cram new words into the finite boundaries of the printed book.... Gove claimed he saved 80 pages in the *Third* by using fewer commas.” (Fatsis 2015)


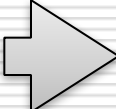
Dictionary conventions: designed to optimize the use of space

băg¹ *n.* **1.** receptacle of flexible material with closable opening at top (esp. w. prefixed word showing contents or purpose; DIPLOMATIC *bag*, GAME¹ *bag*, HAND¹*bag*, KIT¹*bag*, *mailbag*, *travelling-bag*, VANITY *bag*); (w. such prefix understood) particular kind of this; hence ~**FUL**. **2** *n.* **2.** contents of bag; MIXED *bag*; amount of game a sportsman has shot or caught (also fig.) **3.** ~**and baggage**, with all belongings; ~**of bones** lean creature; (whole) ~**of tricks** every... [etc]



- *Concise Oxford Dictionary* 1982
 - “reduced” definitions, abbreviations, cross-references, use of tildes (~) etc

Definition styles in the English tradition: Merriam-Webster's *Collegiate*

- *expectant* characterized by expectation 
- *expectation* the act or state of expecting 
- *expect* [full definition - at last!]
- Features
 - formulaic style: “the act of X-ing”, “characterized by X”
 - recursiveness: full definition for core item, related items refer back to it
 - economical - but tough for users

Distribution of example sentences

- Policy in Macmillan dictionary, print edition (MEDAL)
 - for core vocabulary (top 7500), detailed description, multiple examples to illustrate combinatorial behaviour (syntax, collocation)
 - **but** for infrequent words, minimal treatment, few or no examples
- Driven by space constraints - *not* because under-specifying infrequent words is a good idea

3. Current practice: a complex picture

- ❑ Several models co-exist
 - ❑ print-only, or print-derived but in print *and* digital media
 - ❑ most English dictionaries (mother-tongue or learner's)
 - ❑ print-derived but digital-only
 - ❑ e.g. Wordnik, dictionary.com, MEDAL
 - ❑ from-scratch digital-only
 - ❑ e.g. elexiko (IDS Mannheim), Diccionario de Aprendizaje del Español como lengua Extranjera (DAELE: Pompeu Fabra), Algemeen Nederlands Woordenboek (ANW: INL, Leiden)

The most basic model: print dictionary in digital format

traducción.

(Del lat. *traductiō*, *-ōnis*).

1. f. Acción y efecto de traducir.
2. f. Obra del traductor.
3. f. Interpretación que se da a un texto.
4. f. *Ret.* Figura que consiste en emplear dentro de la cláusula un mismo adjetivo o nombre en distintos casos, géneros o números, o un mismo verbo en distintos modos, tiempos o personas.

~ directa.

1. f. **traducción** que se hace de un idioma extranjero al idioma del traductor.

- *Diccionario de la lengua española* (Real Academia): abbreviations, tildes, recursive and formulaic definitions (“acción de traducir”)...but no hyperlinking

Old and new together: *Wordnik*: web-derived examples, old definitions

expectation

Help support Wordnik by [adopting your favorite word!](#)

[Define](#)

[Relate](#)

[List](#)

[Discuss](#)

[See](#)

[Hear](#)

[Love](#)

Definitions

from The American Heritage® Dictionary of the English Language, 4th Edition

- n.* The act of expecting.
- n.* Eager anticipation: eyes shining with expectation.
- n.* The state of being expected.
- n.* Something expected: a result that did not live up to expectations.
- n.* Prospects, especially of success or gain.
- n.* Statistics The expected value of a random variable.
- n.* Statistics The mean of a random variable.

from Wiktionary, Creative Commons Attribution/Share-Alike License

- n.* The act or state of expecting or looking forward to an event as about to happen.
- n.* That which is expected or looked for.
- n.* The prospect of the future; grounds upon which something excellent is expected to occur; prospect of anything good to come, especially of property or rank.
- n.* The value of any chance (as the prospect of prize or property) which depends

Support



Help Wordnik continue as an ad-free not-for-profit by adopting the word **expectation** [here](#).

Examples

If people want to filibuster so they can debate, then my expectation is they should take the floor, I should tune-in C-Span, and I can hear what their gripe is.

Think Progress » Bayh: 'I Think The Filibuster Absolutely Needs To Be Changed'

Vincent said he hopes Selig will do nothing and let the banishment stand, but my expectation is almost exactly the reverse.

USATODAY.com - Recoanizina 'I'm 14 uears late.' Rose admits he bet

Old and new together: OALD: cross-reference policy

- ❑ Old ones (from printed book): at bottom
- ❑ New ones, added for digital version: at right
- ❑ Do we need both?

  **area** *noun*

BrE /'eəriə/ ; NAmE /'eriə/ 

★ Add to my wordlist

part of place

1  [countable] part of a place, town, etc., or a region of a country or the world

- mountainous/desert areas
- rural/urban/inner-city areas
- There is heavy traffic in the downtown area tonight.
- She knows the **local area** very well.
- John is the London area manager.
- Wreckage from the plane was scattered over a **wide area**.
- The farm and **surrounding area** were flooded.

→ SEE ALSO **catchment area**, **conservation area**, **development area**, **no-go area**

Other results

All matches

penalty area *noun*

goal area *noun*

grey area *noun*

area code *noun*

rest area *noun*

council area *noun*

Broca's area *noun*

service area *noun*

staging area

catchment area *noun*

disaster area *noun*

assisted area *noun*

no-go area *noun*

conservation area *noun*

Old and new together: *Wiktionary*

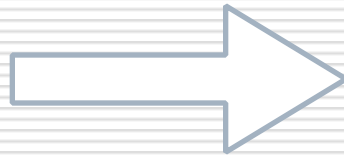
- ❑ Modern methodology: crowdsourcing, web-based
- ❑ Good for terminology: plenty of specialists out there (cf. *Wikipedia*)
- ❑ Weak for everyday vocabulary:
 - ❑ many definitions copied from 1911 edition of Webster
 - ❑ no corpus basis, perpetuates older defining styles
- ❑ “It seems that the web community, while enthusiastically embracing the novelty of online collaboration, propagates the traditional model of lexicographic description” (Lew 2014. 17).

Changes in macrostructure

- ❑ Traditional model
- ❑ Alphabetical order the only navigational aid
 - ❑ an arbitrary system: unrelated words are adjacent
 - ❑ e.g. *redneck, redness, redo, redolent, redoubtable*
- ❑ Phrasal verbs and phrases “nested” under main entries
 - ❑ at **set**, find *set off, set up, set the cat among the pigeons, set sb’s mind at rest* etc

Changes in macrostructure

- Emerging model
- Every lemma separate: alpha order irrelevant
- Related lemmas shown in side panel



- Phrases and phrasal verbs as separate, standalone entries
 - easily found by “bag of words” search
 - e.g. “bolted horse” finds the phrase *close the stable door after the horse has bolted*

Related words

area code NOUN

grey area NOUN

rest area NOUN

safe area NOUN

picnic area NOUN

disaster area NOUN

no-go area NOUN

[more dictionary definitions](#)

Designed for digital: DAELE

- Menu shows senses
- + and - to expand or reduce elements
- click to show conjugation
- corpus examples

poner/se (verbo)

⊕ Conjugar

- ⊕ 1 HACER QUE ESTÉ / ESTAR EN UN LUGAR
- ⊕ 2 HACER QUE ALGO ESTÉ DE CIERTO MODO
- ⊕ 3 EMPEZAR A HACER
- ⊕ 4 DAR, OFRECER
- ⊕ 5 IMPONER
- ⊖ 6 HACER

• **transitivo** Alguien pone una determinada expresión cuando la tiene o la hace:

- *El niño empezó a poner caras divertidas. (SWC)*
- *Puso cara de asombro, levantando las cejas. (SWC)*
- *Traté de superar el asunto y poner buena cara. (SWC)*
- *Ponía mala cara sin venir a cuento. (SWC)*
- *Los niños ponen cara de aburridos y bostezan. (DAV)*

⊕ 7 OPINAR

- ⊕ **poner de mi/tu/su... parte** (locución verbal)
- ⊕ **poner el grito en el cielo** (locución verbal)
- ⊕ **poner la mano en el fuego** (locución verbal)
- ⊕ **poner la mesa** (locución verbal)
- ⊕ **pongamos que +** (frase)

Designed for digital: elexiko (IDS)

- Option to “show” or “hide” examples
- *Explanation* of meaning (not classical definition)
- Tabs give access to other information types (e.g. collocation, constructions, morphology)

elexiko

Beratung 

Lesart: 'Auskunft geben'

 zur Übersichtsseite

Lesarten im Überblick 

Bedeutungs-
erläuterung

Kollo-
kationen

Konstruk-
tionen

Sinnverwandte
Wörter

Gebrauchs-
besonderheiten

Grammatik

 **Erläuterung der Bedeutung / Funktion**

Mit **Beratung** bezeichnet man eine Sprechhandlung, bei der eine Person(engruppe) eine andere Person(engruppe) über einen Sachverhalt informiert und gegebenenfalls Empfehlungen gibt.

 Belege [verbergen x](#)



Nach einem Einkauf in der Spargelstädter Metzgerei kann der Kunde zu Hause im Handumdrehen ein ganzes Menü zaubern. "So wird Fast-Food auf hohem Niveau möglich", sagt Blüm. Auch **Beratung** wird in der Metzgerei großgeschrieben. Bei Laktose-Unverträglichkeit etwa benötigten Verbraucher detaillierte Informationen über die Kost, die sie kaufen. "Wir wissen genau, was wir in unsere Produkte hineingeben, und können daher zuverlässig Auskunft geben", macht der Fleischereimeister deutlich. (Mannheimer Morgen, 30.04.2010, S. 15.)

Transition from print to digital implications

18

Current practice: observations

- ❑ Most commercial dictionaries have a print origin
- ❑ All “aggregators” (e.g. dictionary.com) use older dictionaries as a base
 - ❑ e.g. old editions of *American Heritage Dictionary*
- ❑ “From-scratch” digital dictionaries are a minority
 - ❑ tend to be non-commercial, usually the most innovative
- ❑ A transitional phase
 - ❑ lots of trial-and-error, no clear patterns emerging
 - ❑ few taking full advantage of the new medium

4. Going digital: general principles

- What we need now
- “models for e-dictionaries that focus on critical areas like the data to be included ... the structures to present and accommodate the data, the functions of these dictionaries and the way they should respond to the needs of their target users” (Gouws 2014.157)

Infinite space - so anything goes?

- ❑ John Simpson on the OED (Simpson 2014.21)
- ❑ “On the one hand ... computer storage devices remove size restraints, [so] editors can publish as much associated material as they wish”
- ❑ “On the other hand...if editors were to allow the extent of individual entries to range out of proportion to utility this would result in making the user’s task of interpreting an entry much more difficult.”
- ❑ **Conciseness still a virtue:** avoiding “information overload”

Two key principles still apply

- Consider in terms of
 - “**Analysis**” stage: gathering and analysing data to populate the dictionary database
 - “**Synthesis**” stage: configuring relevant data to create entries for a specific dictionary (cf. Atkins 1993)
- Analysis: must be based on objective evidence of usage
 - frequency, dispersion, longevity
- Synthesis: determined by user-profile
 - who is the user? what are their reference needs and expectations?

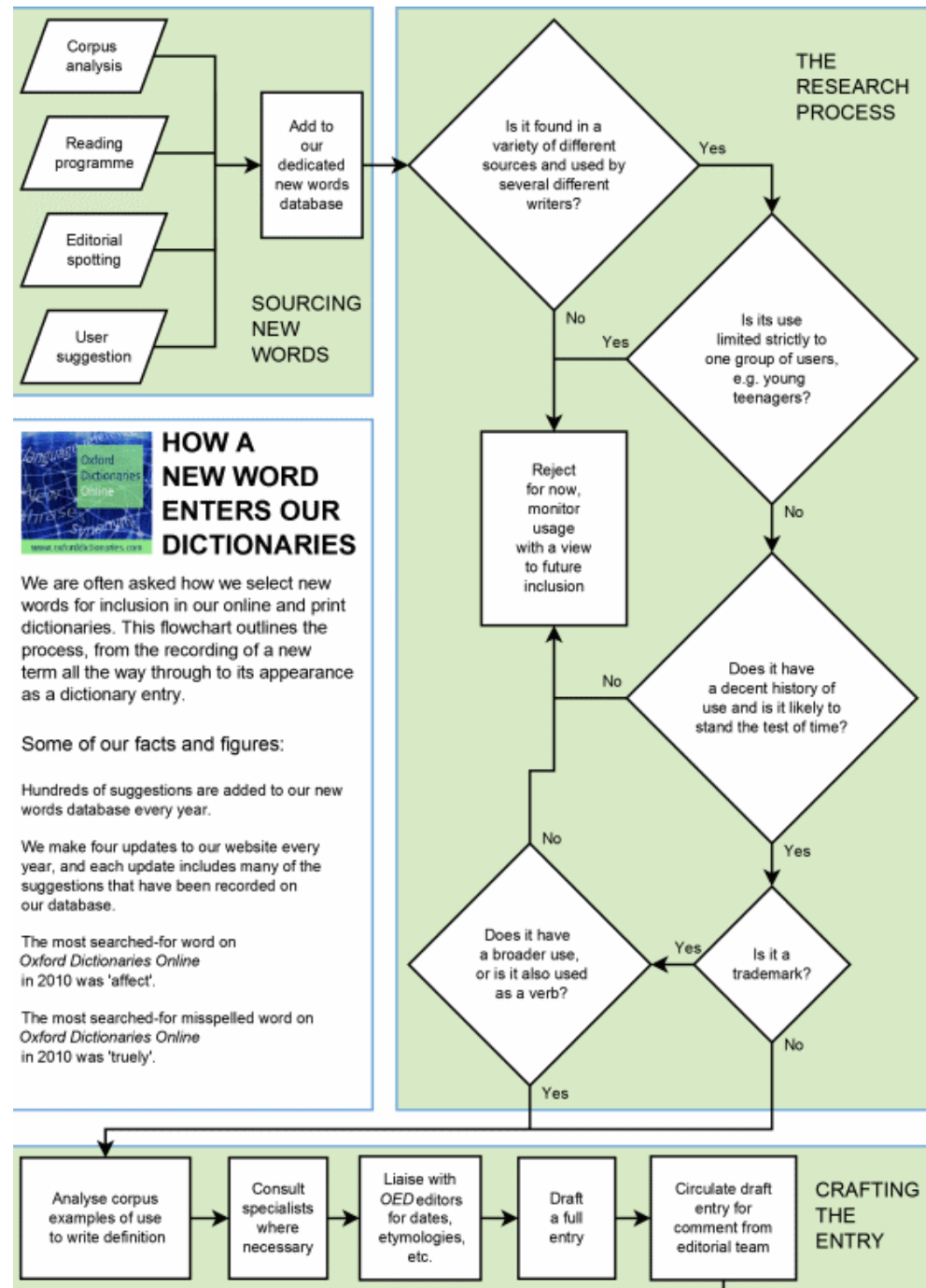
5. Going digital: specific issues

- Inclusion: which words go in the dictionary?
 - in print medium: a critical decision, strict inclusion criteria required
 - in digital media: are these criteria still valid?
- Definitions
 - implications for *wording* and *content*?
 - from static to dynamic microstructure
- Example sentences: how many, where from, ...?
- Linked data: intelligent links within and beyond the dictionary

Inclusion criteria

- The question most asked of a dictionary editor is “How does a word get into the dictionary?”
Peter Sokolowski, Merriam-Webster
- <http://www.merriam-webster.com/video/0032-howaword.htm?&t=1326227263>
- Theoretical point: the lexicon is dynamic
 - An “unbounded set”, novel items always being created
 - For a fuller discussion, see Hanks 2013. 25-64

- Inclusion: the traditional model (OUP's flowchart)
- Strict entry criteria
- Every word has to earn its place (since space is limited)
- Hence: dictionary's role as “gatekeeper”



Filling in gaps

- E.g. MEDAL (in print form) *selectively* includes chemical elements - on basis of salience
 - included : *potassium, calcium, hydrogen*
 - excluded: *flevorium, meitnerium, niobium*
- In digital edition, we fill in the gaps: *all* elements included
- Both policies defensible
 - in print mode, assess relative “value” of each
 - in digital mode, common sense: no reason *not* to complete the set

Are hashtags words?

- ❑ Readers' poll for Collins: which word do you want to add to the dictionary?
 - ❑ *#nomakeupselfie* came fourth!
- ❑ American Dialect Society “Word of the Year” 2014
 - ❑ *#blacklivesmatter*
 - ❑ coined in 2012, after Trayvon Martin killed in U.S.
 - ❑ discussed at www.macmillandictionaryblog.com/blacklivesmatter-and-words-of-the-year
- ❑ But: transparent, not used in “normal” discourse - needed in a dictionary?

Inclusion criteria: turn the question around

- ❑ **Formerly:** does this word pass our tests? Are there good enough reasons to *include* it?
- ❑ **Now:** are there good reasons to *exclude* it?
- ❑ Possible “exclusion criteria”
 - ❑ unlikely to benefit our users (e.g. highly technical expert-to-expert terminology)
 - ❑ likely to be ephemeral (not always easy to judge)
 - ❑ extremely parochial: e.g. very local dialect words

Definitions: wording

- ❑ Defining Vocabularies (DV)
- ❑ Most (English) learner's dictionaries use a controlled DV
 - ❑ limited set of high-frequency words (typically 2000-3000) to be used in all definitions
 - ❑ goal is comprehensibility for users with limited proficiency (as low as B1)
 - ❑ some research supports this

Defining vocabulary

- But what if every word in the definition is hyperlinked?
 - mouseover at “lava” links to entry for *lava*

erupt - definition



VERB [INTRANSITIVE]



Pronunciation

/ɪˈrʌpt/

Word Forms

- 1 if a volcano erupts, it explodes inside and flames, rocks, and lava come out of the top

T Thesaurus entry for this meaning of erupt

lava

Definitions: content

- Opportunities for multiple versions
- e.g.1 decoding and encoding options
 - decoding: minimal needed for the task
 - encoding: fuller information, both semantic and pragmatic
- e.g.2 layperson and technical options
 - lay and technical definitions of *fermentation*
 - plus “defining examples” from corpus
 - “Fermentation refers to any process by which large organic molecules are broken down to simpler molecules as the result of ...”

Definitions: latest from Merriam-Webster

- ❑ Two layers of definition
- ❑ top ones more accessible
- ❑ “full” definitions less usable!

expectation

noun | ex·pec·ta·tion | \ˈek·spek-ˈtā-shən, ik-\

SAVE PO



 Cite!

 Share

 +1

: a belief that something will happen or is likely to happen

: a feeling or belief about how successful, good, etc., someone or something will be



Top 10 words to win with in SCRABBLE® »

Full Definition of EXPECTATION

1 : the act or state of **expecting** : **ANTICIPATION** <in *expectation* of what would happen>

2 a : something **expected** <not up to *expectations*> <*expectations* for an economic recovery>

b : basis for expecting : **ASSURANCE** <they have every *expectation* of success>

Example sentences

- Recent development: the “extra examples” option
 - “Click for more examples”
 - Typically from corpus or web or social media
 - Increasingly common feature
- Mapping examples to “correct” word-sense: labour-intensive if manual, difficult/unreliable if automated

“More example sentences” in ODO

party¹



See definition in [Oxford Advanced Learner's Dictionary](#)

★ Top 1000 frequently used words

Line breaks: party

Pronunciation: /'pa:ti/

noun (plural parties)

- 1 A social **gathering** of **invited guests**, typically involving eating, **drinking**, and entertainment:
'an **engagement party**'

[MORE EXAMPLE SENTENCES](#)

[SYNONYMS](#)

- 2 A formally **constituted** political group that **contests** elections and attempts to form or take part in a government:

- 1 A social **gathering** of **invited guests**, typically involving eating, **drinking**, and entertainment:
'an **engagement party**'

[MORE EXAMPLE SENTENCES](#)

'Sometimes, caterers serving at parties and social gatherings order large quantities.'

'The hotel staff encourages the use of this area for social gatherings and parties.'

'A buffet of finger foods is the perfect way to serve guests at an anniversary party or wedding reception.'

[GET MORE EXAMPLES](#)

Examples: links to corpus data

- DWDS (German)
- also found in
 - ANW (Dutch)
 - Den Danske Ordbog (Danish)

DWDS-Wörterbuch

Hausarrest Detailansicht

mask.; -s/-es; -e

Aussprache: ▶

◆ Freiheitsentzug, bei dem es dem Betroffenen verboten ist, das eigene Haus zu verlassen ↓

dwdsweb-0.3.41

Kernkorpus 20 (eingeschränkte Version)

Im Kernkorpus wird zwischen Groß- und Kleinschreibung unterschieden.

Treffer: 4 1

- 1 ...998 in Rom auftaucht und unter **Hausarrest** gestellt wird. Überall in E...
- 2 ...ten Vertreter der Nagas, unter **Hausarrest** gestellt; zweimal täglich muß ...
- 3 ...tt der Staatslimousine, in den **Hausarrest** verbracht wurden. Das war, ...
- 4 ...geprügelt werden wollte, mußte **Hausarrest** wahren oder auswandern, nicht ...

nach Anmeldung 71 von 97 Treffern anzeigbar DDC-Query | Darstellung | Suchfilter

ZEIT & ZEIT online

Treffer: 738 1 2 3 4 >>|

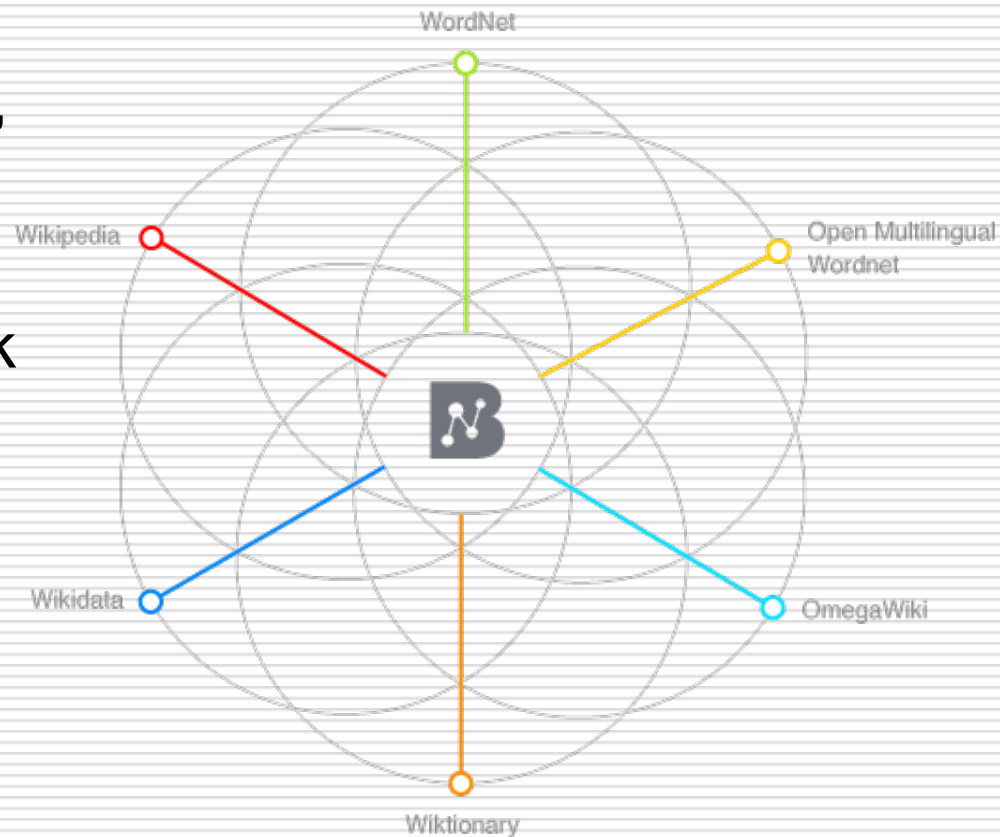
- 1 ...ilitärjunta die Auflagen ihres **Hausarrests** verletzt hat. UN-Generalsek...
- 2 ...zer Laufhund nach fünf Monaten **Hausarrest** . © Self contained ist unse...
- 3 ...gnis, eine Tochter steht unter **Hausarrest** . Die eifache Mutter, die ...
- 4 ...rägerin, die seit Jahren unter **Hausarrest** steht "als wenn das alles wäre.
- 5 ...damals 85-jährigen umgehend in **Hausarrest** umgewandelt wurde. Denn das

Linked Open Data

- Huge research area
 - A “web of data” where “links are at the level of data, as a counterpart to the traditional Web, in which links are established at the level of documents”
 - linking lexicons, ontologies, corpora, encyclopedias...
 - see Gracia 2015
 - so far, underexploited for lexicography, but great potential

Linked data - an example: Babelnet

- “A multilingual encyclopedic dictionary, with lexicographic and encyclopedic coverage, **and** a semantic network which connects concepts and named entities”
- <http://babelnet.org/about>



6. Conclusions: redefining “dictionary”

diction /'dɪkʃ(ə)n/ noun [U] **1** the way that you pronounce words, especially whether or not you speak or sing clearly **2** the choice of words used in a speech or piece of writing

dictionary /'dɪkʃən(ə)ri/ noun [C] ★★

1 a book that gives a list of words in alphabetical order and explains what they mean: *a dictionary of the English language* **1a.** a book that lists words in one language and gives translations in another: *a German-English dictionary*

2 a book about a particular subject that gives an alphabetical list of words, phrases, or names with information about them: *a dictionary of art/music*

dictum /'dɪktəm/ noun [C] an expression or statement that people often repeat because it says something interesting or wise about a subject

Definition of “dictionary” in 2007 (MEDAL2, paper)

Redefining “dictionary” (2)

dictionary - definition



NOUN [COUNTABLE]



Pronunciation

/ˈdɪkʃən(ə)ri/

Word Forms

- 1 a reference resource which provides information about words and their meanings, uses, and pronunciations. A dictionary may be published as a printed book, or as a digital product such as a website or app, and it may be monolingual, bilingual, or multilingual.

a dictionary of the English language

an English-Chinese dictionary

Definition of “dictionary” in 2015 (MEDAL online)

Redefining “dictionary” — again

- ❑ A one-stop reference resource
 - ❑ dictionary **and** thesaurus, monolingual **and** multilingual
 - ❑ links to corpora and other data from the Web
 - ❑ language resources e.g.
 - ❑ blog, grammar tips, “what’s the difference?”, teaching materials
 - ❑ games, competitions, quizzes, videos ...
 - ❑ User-generated content (UGC)

Exploiting digital media: work in progress

- ❑ Traditional dictionary policies are familiar to users, and have a long history - digital dictionaries still relatively new
- ❑ Current dictionaries in an interim state - older policies and structures co-exist with innovations (in varying degrees)
- ❑ What is needed now:
 - ❑ more user-research on online dictionaries
 - ❑ exploit opportunities of digital media, e.g. taking advantage of universal (therefore familiar) web conventions
 - ❑ maintain basic principles: evidence-based, sensitive to user-profile

-
- Thank you
 - Ngiyabonga