

# It works in practice – but will it work in theory?

---

*The uneasy relationship between  
lexicography and matters theoretical*

**Michael Rundell**  
**Lexicography MasterClass**  
**& Macmillan Dictionaries**



# Zgusta's dilemma

---

- 'Every lexicographer knows that rosary of agonizing choices that must be taken every time when a decision has to be made and there is no time to do what would be the only reasonable thing to do, namely to research the problem for a year or two ...
- ...trying to find the underlying regularity or rule, in ... chaos or randomness' (Zgusta 1992. 91-2).

# Describing language

---

## □ Lexicographers' goals

- reduce reliance on intuition, subjective judgments
- replace (where possible) with objective criteria
- completeness: ensure nothing important is omitted
- system: internally-consistent dictionaries

## □ How?

- driven by language data
- perceiving order in 'chaos or randomness'
- → **role of theory**

# But what kind of theory?

---

1. Metalexicography
  - Shcherba
  - Wiegand
2. Function Theory: the 'Aarhus School'
3. Linguistic theories
4. The world we live in now: e-lexicography
  - are existing theories still valid? do we need new theories?

# What is a 'theory' anyway?

---

**theory** Pronunciation: /'θiəri/

Translate **theory** into [French](#) | [into German](#) | [into Italian](#) | [into Spanish](#)

Definition of **theory**

*noun (plural theories)*

a supposition or a system of ideas intended to explain something, especially one based on general principles independent of the thing to be explained:

*Darwin's theory of evolution*

- a set of principles on which the practice of an activity is based:

*a theory of education*

*[mass noun]:*

*music theory*

**Source: ODE**

# 'Theory' in the language data

**theory** (*noun*) LEXMCI freq = **159149** (92.5 per million)

<u>object of</u>	<u>36199</u>	<u>1.9</u>	<u>PP X</u>	<u>36927</u>	<u>AJ premod</u>	<u>52185</u>	<u>2.9</u>	<u>and or</u>	<u>39496</u>	<u>1.6</u>	
ground	<u>283</u>	7.63	PP of-i	<u>25872</u>	3.6	evolutionary	<u>844</u>	8.89	practice	<u>5693</u>	8.01
disprove	<u>154</u>	7.0	PP in-i	<u>2366</u>	0.6	feminist	<u>443</u>	7.91	theory	<u>1583</u>	7.44
test	<u>606</u>	6.95	PP to-i	<u>1388</u>	0.6	Marxist	<u>424</u>	7.83	philosophy	<u>399</u>	6.71
underlie	<u>327</u>	6.92	PP about-i	<u>1206</u>	5.0	literary	<u>555</u>	7.59	methodology	<u>294</u>	6.54
formulate	<u>199</u>	6.79	PP on-i	<u>998</u>	0.6	scientific	<u>974</u>	7.56	concept	<u>684</u>	6.54
apply	<u>1016</u>	6.61	PP for-i	<u>934</u>	0.5	mathematical	<u>465</u>	7.49	relativity	<u>100</u>	6.2
integrate	<u>204</u>	6.11	PP with-i	<u>752</u>	0.6	psychological	<u>503</u>	7.43	geometry	<u>135</u>	6.18
advance	<u>146</u>	6.02	PP into-i	<u>685</u>	2.3	critical	<u>906</u>	7.35	algebra	<u>96</u>	5.98
develop	<u>1602</u>	5.96	PP as-i	<u>487</u>	0.7	classical	<u>469</u>	7.31	logic	<u>162</u>	5.96
compete	<u>142</u>	5.92	PP behind-i	<u>467</u>	20.8	sociological	<u>296</u>	7.31	mechanic	<u>114</u>	5.83
propound	<u>65</u>	5.83	PP from-i	<u>380</u>	0.4	psychoanalytic	<u>245</u>	7.21	experiment	<u>273</u>	5.8
learn	<u>922</u>	5.81	PP at-i	<u>274</u>	0.3	linguistic	<u>325</u>	7.2	economics	<u>119</u>	5.67

# ‘Lexicographic theory’?

---

- ‘Strong’ sense: attempts to explain naturally-occurring phenomena; falsifiable; predictive:
  - Darwinism, Standard Model (particle physics), **linguistics**
- ‘Weak’ sense: principles guiding a practical activity
- ‘Theoretical lexicography’: an oxymoron
- (‘Practical lexicography’: a tautology)
- ‘Lexicographic theory’ – not *inherently* wrong

# An aside: dictionary-user research

---

- By publishers: market research
  - usually unpublished
  - influential: e.g. dropping of complex syntax codes in LDOCE and OALD in 1980s
- By academics: studies of user behaviour
  - a practical activity – in a theoretical framework
  - huge body of work, role of Euralex
  - taken seriously by dictionary makers
- Its relevance is not in doubt: not discussed here



# Metalexicography: (1) L.V.Shcherba

---

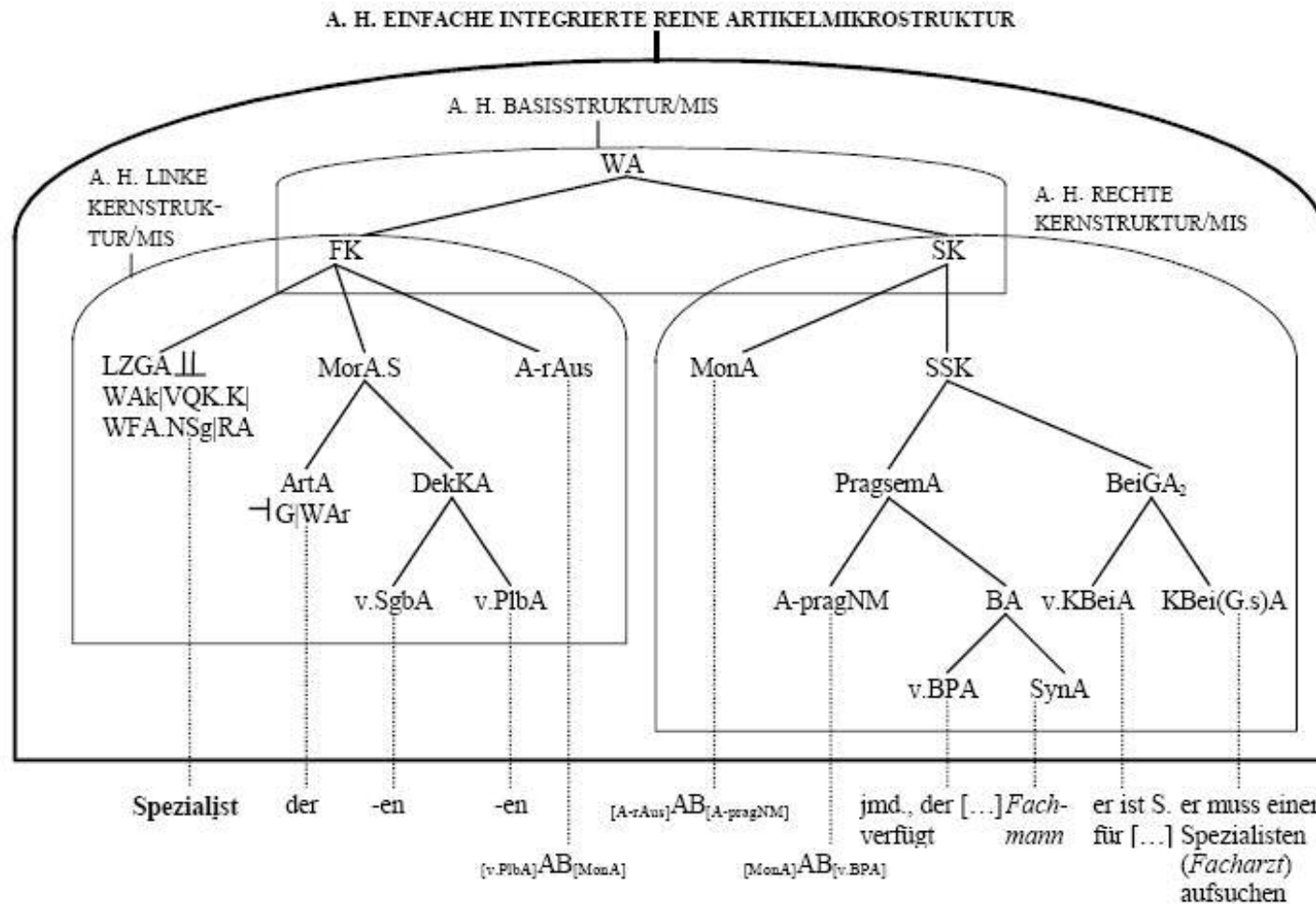
- 'Towards a general theory of lexicography' (1940)
- Six binary 'oppositions', e.g.
  - academic vs informative dictionary
  - defining vs translating
  - ordinary vs ideological
  - dictionary vs encyclopedia
- Raises many issues – provides few solutions
- **A typology is not a theory** – even in the 'weak' sense

# Metalexicography: (2) H.E.Wiegand

---

- Wiegand 2010 (Lexikos 20)
- ‘a contribution to the theory of dictionary structures’
- 20 Keywords (English version), including:
  - hierarchical architectonically enriched article  
microstructure, hierarchical hybrid deep double gloss-  
conditioned item structure, hierarchical hybrid shallow  
double gloss-conditioned item, non-functional-positional  
segmentation

# From Wiegand 2010.507

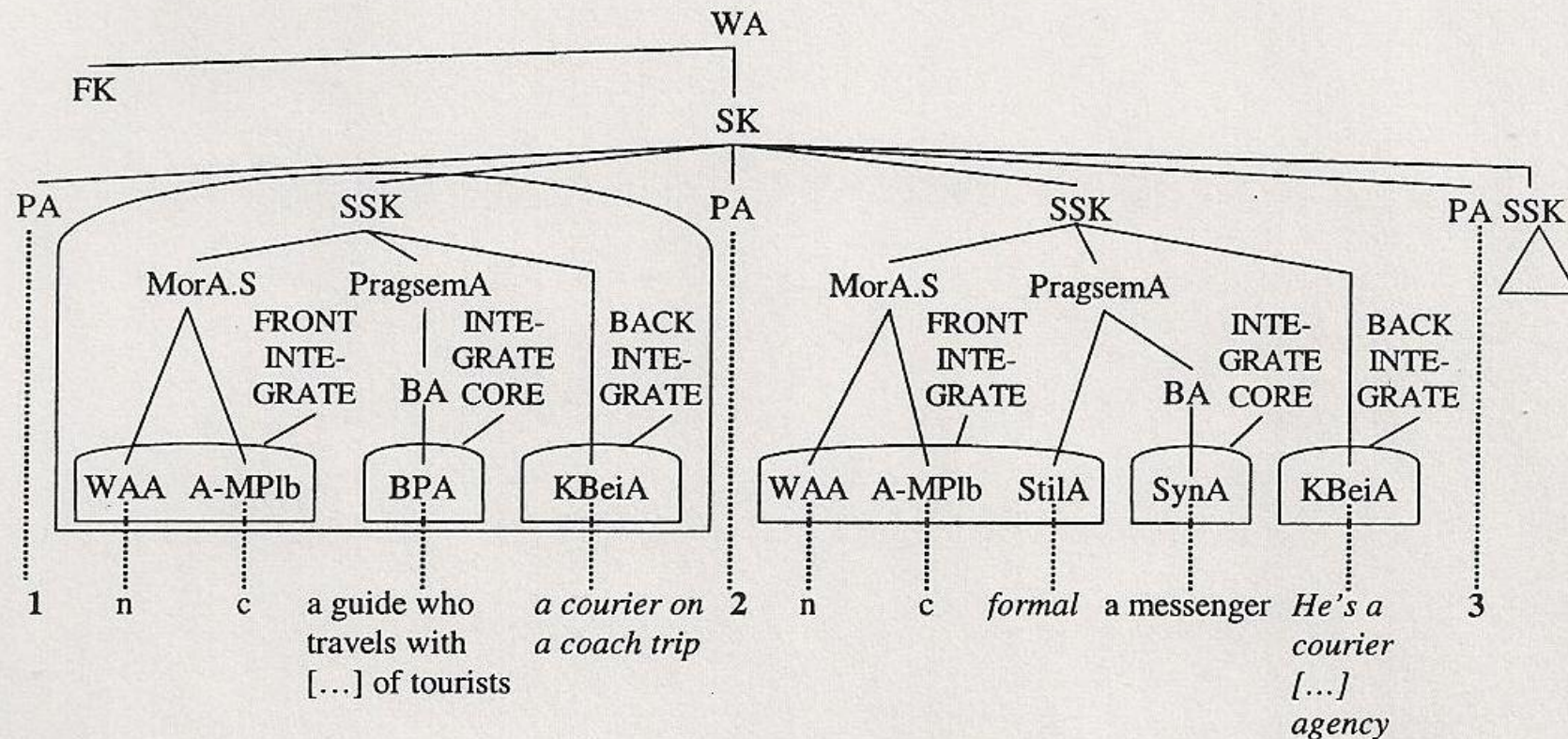


# Wiegand's contribution

---

- Dominant figure in the field: prolific output, enormous range, including
  - dictionary-user research
  - philosophical speculation
  - defining policy
  - **micro- and macro-structural components**

# A simple entry: *courier* (1999.210)



**Fig. 8:** Partially visualized and annotated structural graph of the simple integrated hierarchical microstructure of  $DA_6$ ; new abbreviations: MorA.S = morphology item for nouns, WAA = item giving the word class, A-MPlb = item concerning the possibility of plural formation, StilA = item giving the level of style, SynA = item giving the synonym

# Commentary

---

- A simple entry
  - 10 datatypes (DANTE database worked with 92)
- ‘The partial string WAA < A-MPIb forms the front integrate (cf. fig 8), i.e. the partial string which always immediately follows the item giving polysemy (PA) and which is always positioned immediately before the integrate core’ (Wiegand 1999. 211).

# A lexicographic translation

---

- ‘The grammar code *follows* the sense number and *precedes* the core of the entry (definition, example etc)’
- Expressed as a DTD

```
<!ELEMENT DictEntry (HWDGroup|LUCont)>
```

```
<!ELEMENT HWDGroup (HWD | FORM)
```

```
<!ELEMENT LUCont (POS | GRAM | REG | SYN | STYLE | MeaningGp)
```

```
<!ELEMENT MeaningGp (DEF | EX)
```

# Observations

---

- Arcane nomenclature, granular, complex
- **Who is it *for*?**
- ‘As a lexical semanticist, I confess I found a lot of this heavy going ... and it was not obvious to me as an outsider that it would have much practical import. I suspect the real intended audience is other metalexigraphers’  
Cruse 2001.142



# Urdang's query

---

- 'I was completely confused by a string of entries in the Ds, namely, *diaconnotative information*, *diaconnotative markedness*, *diaevaluative information*, *diaevaluative markedness*, *diafrequential information*, *diafrequential markedness*, and so on, through *diaintegrative . . . . .*, *dianormative . . .*, *diastratic . . .*, *diatechnical . . .*, *diatopic(al) . . . .* **It seemed odd that such terms of art in a field in which I have lived and worked for many years would have eluded me.**
- Review of Hartmann & James, *Dictionary of Lexicography*

# Metalexicography: some conclusions

---

- Descriptive: the dictionary as an object of study: a legitimate intellectual exercise
- Benefits of a stable nomenclature ... within reason
- Applications for working lexicographers?
  - entry structures...yes, but *this is what lexicographers do*, and have always done
- If lexicographers don't pay attention, not through antipathy to theory *per se*

# The 'Aarhus School'

---

- ❑ Bergenholtz, Tarp and colleagues
- ❑ 'Theory of lexicographical functions', aka 'function theory'
- ❑ In development since 1990s, extensive literature
- ❑ For a fine-grained exposition, see 'General theory for learner's dictionaries' (Tarp 2008.125-171)

# Focus on users and situations of use

---

- In planning a dictionary, create ‘a profile of the intended **user group** and a typology of the **user situations** where problems or needs may pop up that can be solved by providing lexicographic data’ (Bergenholtz & Tarp 2003.173)
- ‘It is a dictionary’s functions that ...determine which data it should contain and how this data should be structured and made accessible’ (Tarp 2008.168)

# Who could disagree? But...

---

## □ Claims of originality

- ‘it was not until the appearance of the modern theory of lexicographic functions that a theory was developed that takes the users, the user needs and the user situations as the starting point for all lexicographic theory and practice’ (Bergenholtz & Tarp 2003.172).

## □ What function theory fails to address

## □ A worrying hostility towards linguistics

- ‘linguistic colonialism’ (Fuertes-Olivera & Bergenholtz 2011b: 3)

# An original theory?

---

- 'If [a] dictionary is designed for a special class of users, their special needs must be taken into consideration' (Hornby 1965: 104).
- 'Dictionaries should be designed with a special set of users in mind and for their specific needs' (Householder & Saporta 1962)
- ...not to mention Robert Cawdrey (1604)...

# ‘Hard usual English words...’

---

- With the interpretation thereof by *plaine English words, gathered for the benefit & helpe of Ladies, Gentlewomen, or any other vnskilfull persons.* Whereby they may the more easilie and better vnderstand many hard English wordes, which they shall heare or read in Scriptures, Sermons, or elsewhere, and also be made able to vse the same aptly themselues.

# An incomplete theory

---

- ❑ Deals almost exclusively with delivering information to the user ...
- ❑ ‘...[does not] discuss the risk of delivering false or misleading information. They proceed as if the truth were known and the database contained all and only correct material. Would that it were so!’ (Kilgarriff 2012: 28)
- ❑ **Concerned with ‘synthesis’, not ‘analysis’**



# A bizarre attitude to language data

---

- Participants (in recent FT Symposium) ‘agreed on a rethinking of some of the approaches commonly used for dictionary-making. **One of them is the role of corpora**’
- We are ‘confronted with the sad truth that lexicographers are **required to adapt their work** and their data selection **...to the results generated by the computer**’ (Samaniego Fernandez & Pérez Cabello de Alba 2011.309)

# Linguistic theory and lexicography

---

- Compare:
- (Meta)lexicographic theories
  - developed specifically to inform dictionary-making
- Linguistic theories
  - independent of lexicography
  - *theory* in ‘strong’ sense: intended to explain naturally-occurring phenomena

# Why consult linguists?

---

- Back to lexicographers' goals:
  - **relevance**: what is worth accounting for?
  - **completeness**: nothing relevant unaccounted for
  - **consistency**: similar linguistic facts treated in same way
  - **objectivity**: judgments guided by robust policy (not ad hoc, subjective)
- 'A large proportion of the decisions made by the lexicographer are linguistic decisions' (Atkins 1993.5)

# An entry for *shoot*: what is relevant ?

---

- *She shot him in the leg*
- *She shot him in the kitchen*
- In Frame Semantic terms:
  - FEs are ‘roughly, the things worth talking about when a frame is evoked’ (Fillmore)
  - Here: *in the leg* instantiates the FE ‘subregion’
- → Completeness: other verbs that behave like this (*bash, punch, whack...*)
- → Consistency, objectivity, robust style policies

# Theories with specific relevance (1)

---

## □ Prototype theory

- ‘Well suited as a theoretical basis for [lexicography], since it accurately models the kind of semantic phenomena that lexicographers have to face up to’ (Geeraerts 1990: 210)

## □ Informs (inter alia)

- WSD: ‘preferences’ not ‘rules’ (cf. Hanks TNE)
- Defining policy: goal is ‘typification’
- Entry structures (e.g. ODE)

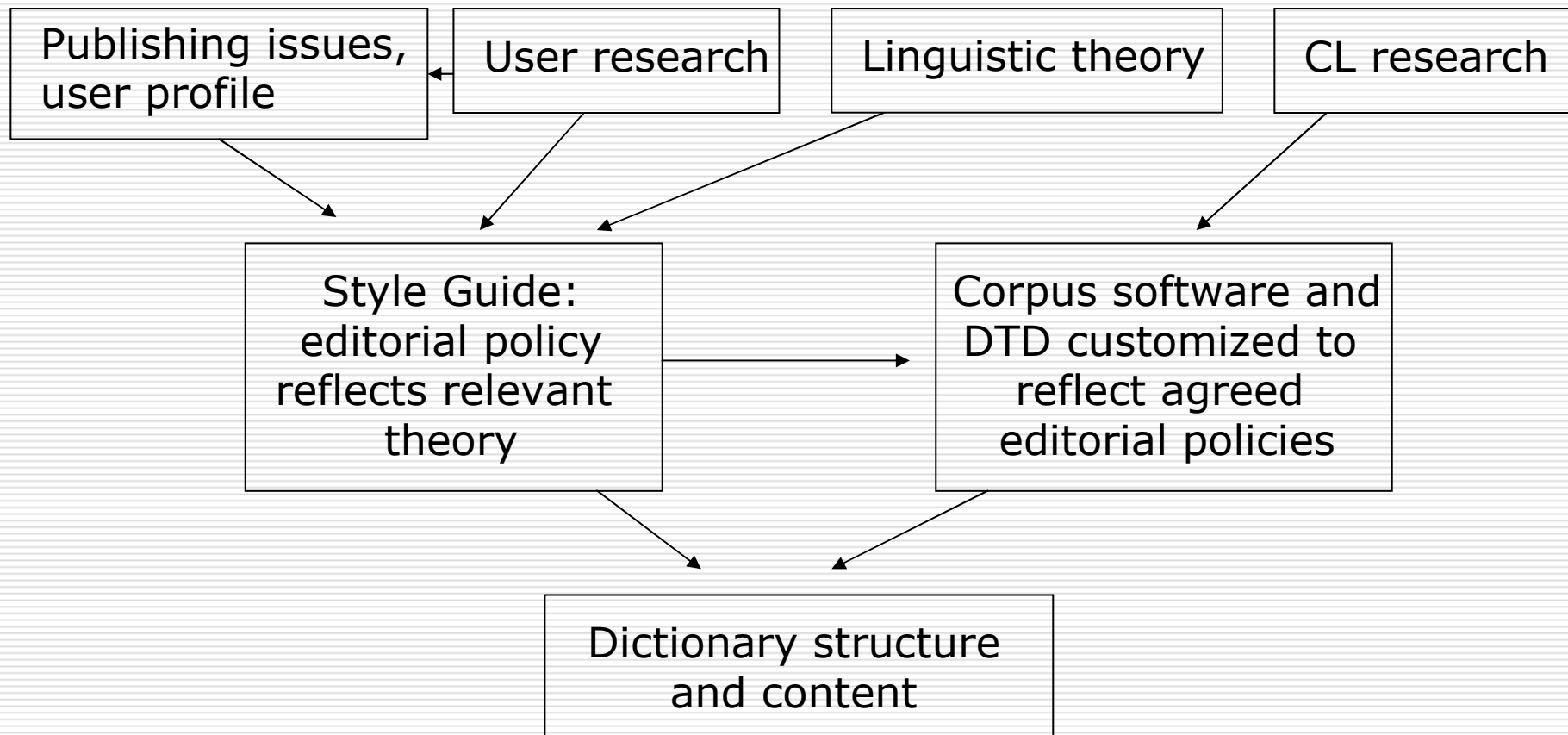
# Theories with specific relevance (2)

---

- Regular polysemy (Apresjan 1974)
  - Members of some semantic sets behave in predictably similar ways, e.g.
    - container and contents: *Three bottles in the dustbin* vs. *We drank three bottles of wine.*
- Allows for ‘proforma’ entries (aka ‘template entries’, Atkins & Rundell 2008. 123-128)
- Applied: OHFD, MEDAL, DANTE (68 proformas)
- Promotes goals of completeness, consistency

# Theoretical ideas, practical outcomes

---



# Automation: streamlining the lexicographic process

---

- Lexicography is labour-intensive
- Which parts can be automated? E.g....
  - example finding (GDEX): requires a 'theory'
    - what makes a 'good' example?
  - WSD/definition writing: needs a theory of how meanings are created, contextually signalled
- You can't automate a process unless you understand it: role of theory
  - first identify systems → then teach machines



# Linguistic theory: conclusions

---

- ❑ ‘Sound lexicography can only be based on sound linguistic theory’ (Apresjan 2002.91)
- ❑ More theory please!
  - e.g. register labelling: needs robust theory
    - ❑ what counts as ‘formal’, ‘informal’ etc?
    - ❑ are these categories even useful?
- ❑ Lexicographers’ autonomy, e.g.
  - polysemy/homonymy distinction: theoretically valid, used in historical dictionaries...but ignored in many
- ❑ Eclectic, yes (Swanepoel); anti-theoretical, no

# The world we live in now

---

- “I have selected a container-grown **Fagus sylvatica Purpurea** from the nursery that supplies the City Council with trees”
- What do I do:
  - Consult a Latin dictionary?
    - ‘*fagus* (from Greek) a beech-tree’
  - Consult a dictionary of botanical terms?
- Neither!

# Implications of e-Lexicography – practical and theoretical

---

- Second major ‘revolution’ in last 30 years
  - ‘Corpus revolution’:
    - profoundly affects lexicographers’ working methods
    - leads to new understanding of language
    - → better dictionaries...but still dictionaries
  - ‘Digital revolution’
    - still unfolding: outcomes hard to predict
    - profoundly affects users’ behaviour: we (producers) no longer have control

# What has changed?

## (1) The obvious things

---

### □ Unlimited space

- dictionary conventions, defining styles : driven by need to maximize scarce space
- inclusion criteria: do old criteria have any validity?

### □ 'Dynamic potential of electronic displays' (Lew)

- multimedia
- entry structures, dynamic sense order (e.g. Kosem)
- data visualization
- hyperlinking

# What has changed?

## (2) Crowd-sourcing/UGC

---

- ❑ Wiktionary, Open Dictionary (Macmillan), Urban Dictionary, Eijiro Pro, user forums etc
- ❑ Good for 'long tail': neologisms, regional varieties, **terminology** (mono- and multilingual)
  - ❑ 'Each contributor has a certain field of expertise... [this] fosters the encoding of a vast amount of domain-specific knowledge' (Meyer & Gurevych)
  - ❑ may make possible what was formerly impracticable
- ❑ Some concerns (accuracy, consistency of approach, access to data): manageable?

# What has changed?

## (3) The 'self-updating dictionary'

---

- New dictionary entries created by software
  - detect emerging words/senses to be added
  - using corpus data, thesaurus features, example-finding tools etc, create and populate entries automatically
  - using parallel corpora, bilingualize
- Not there yet ... but in development

# What has changed?

## (4) The 'adaptive dictionary'

---

- 'Systems adaptively select and prioritise the items which are most relevant to their users' (Kwary 2012.35)
  - 'adaptively' the key word: as users' needs change, dictionary continually resets, reconfigures
  - big Web research area
  - (compare: '**adaptable**' dictionary': user specifies own requirements, proficiency etc)

# What has changed?

## (5) The 'disappearing dictionary'

---

- Analogy: 'Ultimately the word *computer* itself will disappear from the English language' (Michio Kaku, *Physics of the Future*, 2011)
  - Because computers are everywhere, in everything
- Similarly dictionaries?
  - embedded in other devices (e.g. Kindle, websites)
  - translation memory, e.g. [www.linguee.com](http://www.linguee.com)
  - text remediation tools (e.g. Prinsloo, Heid et al. 2011)
- If users' needs can be met by other means...?



# What has changed?

## (6) What users do

---

- 'People typically consult maps, encyclopedias and dictionaries while they are doing something else' (Nesi, in press)
- MEDAL's experience
  - 66% of visitors arrive via a search engine; of these SE searches, only 17% are 'branded'
- It's all about '**search**', not dictionaries
  - humans: no fixed pattern, skills of digital natives
  - software: improving search algorithms, SEO
- **We are not in control**

# Where does this leave our theories?

---

- Risk of entropy
  - ‘a jamboree of all these dreams’ (de Schryver 2003.188)
  - so we still need guiding principles
- The key ones still apply
  - true to the data: description reflects evidence of usage
  - optimised to needs and capabilities of users
  - goals of completeness, self-consistency
- How to facilitate this?

# Don't resist change – manage it

---

- ❑ *If* users land on your site...ensure content conforms to same principles. But how?
- ❑ Own content and structure
  - adapt to/exploit new media, input from information science, emerging linguistic theories etc
- ❑ Autogenerated content, adaptive technologies:
  - lexicographers' input at development stage
- ❑ Crowd-sourced content
  - entry templates/protocols, give access to data etc
- ❑ Collaborative model: part of a larger enterprise

# It works in practice but will it work in theory?

---

- Different academic/intellectual traditions (British, continental Europe)
  - ‘If lexicography were denied any form of being a science, linguistics would remain a discipline at our universities, but lexicography would not. We would not be able to do any dictionary work at universities’ Bergenholtz & Bergenholtz 2011.189
  - British perspective: retrofitting theory to practice, to confer academic respectability?

# What would Hornby do?

---

- 'A very practical man' (Quirk)
- Influences are both practical and theoretical
  - linguistic theory (Palmer, IRET etc)
  - language teaching experience
  - → lexicography: ISED
- An innovator