

Collect and correct

– on gathering and assessing terminological resources



Henrik Nilsson
Terminologikum TNC
(Swedish Centre for Terminology TNC)
International Seminar, Bangor (Wales)
4 September 2017

© Terminologikum TNC

Agenda

- TNC, terminology work in Sweden, the national termbank Rikstermbanken
- Collecting terminological resources
 - methodology
 - state of existing resources
 - classification
- Correcting terminological resources
 - recurring “mistakes”
- Future developments (crowdsourcing, linked open data)



- Founded by engineers in 1941 as Tekniska nomenklaturcentralen (Swedish Centre for Technical Terminology)
- Private company with varying state grants
- Restructured as Terminologikum TNC in 2000
- 5 people employed today
- 105 publications, cd-rom Termdok (1988)
- national termbank (Rikstermbanken) (2009)

Terminologacentrum TNC (Swedish Centre for Terminology)

- terminological services and support to authorities, organizations and enterprises
- terminological products, e.g. terminological glossaries and databases, manuals for technical writing
- **collection, processing and dissemination of terminology of specialised subject fields**
- terminological reviews of standards and other documents containing terminology
- lectures and courses on the principles and methods of terminology work and technical writing
- co-operation with other language institutions on a national level and with terminological institutions on an international level
- **maintains the national termbank, Rikstermbanken**

EAFT: The Board

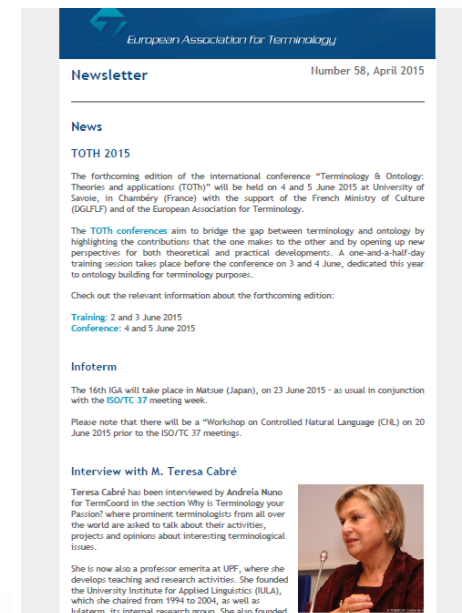
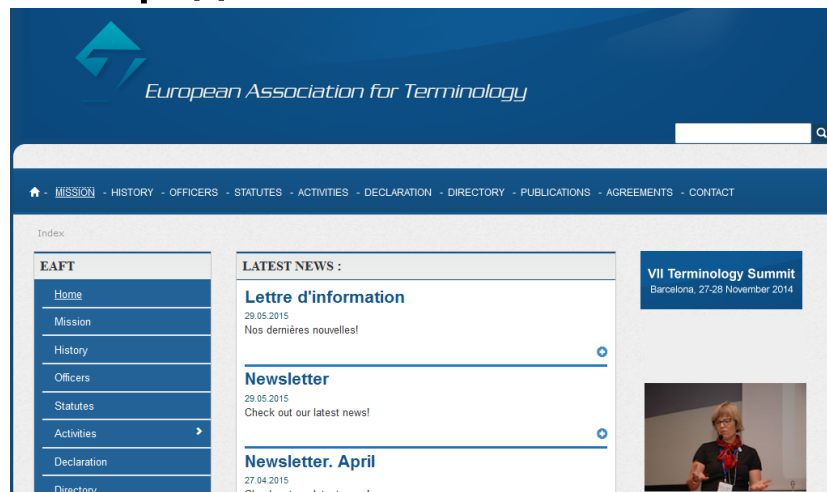


From left:

- **Henrik Nilsson**, president (Terminologikum TNC, Sweden)
- **Anca-Marina Velicu** (University of Bucharest, Romania)
- **Tegau Andrews** (Bangor University, Wales)
- **Eszter Papp**, vice president (University of Károli Gáspár, Hungary)
- **Imanol Urbieto** (Uzei, Basque country)
- **Sandra Cuadrado**, secretary and treasurer (Termcat, Catalonia)
- **Ana Ostroški** (Institute of Croatian Language and Linguistics, Croatia)

Activities

- Terminology Summit (latest: 2016, Luxembourg)
- ITA (International Terminology Awards)
- Brussels Declaration
- Agreements (Nordterm, ELRA, LTT, Infoterm)
- Newsletter (EAFT Newsletter)
- SIG:s (special interest groups)
- <http://eaft-aet.net>



www.rikstermbanken.se

[Ingångssidan](#)[Enkel sökning](#)[Utökad sökning](#)[Källor](#)[Senaste nytt](#)[Frågor och svar](#)[Hjälp](#)

Välkommen till Rikstermbanken

Sveriges nationella termbank

Sök i Rikstermbanken

☒ enbart bland termer ☐ även i all annan text

Nya och reviderade källor läggs in successivt, se [Senaste nytt om Rikstermbanken](#).

Hittar du inte vad du söker eller har du termproblem du vill ha hjälp med? Gå med i Terminologifrämjandet och använd dig av TNCs termtjänst. [Läs mer](#).

European "national" termbanks

Stofnun Árna Magnússonar
í íslenskum fræðum, Iceland:
Orðabanki

Norway, :
**Termportalen,
Snorre**

TNC, Sweden:
Rikstermbanken

TSK, Finland:
Vetenskapstermbanken, TEPA, Valter

Eter, Estonia:
ESTERM

Latvia:
AkadTerm

LKI, Lithuania:
Terminų bankas

Denmark, :
(DTB)

NL-Term,
Nederlanderna:
Nedterm

Foras na Gaeilge,
Ireland:
Téarma.ie

Wales:
**National
Terminology
Portal**

Société française de
terminologie, Confédération suisse
France:
FranceTerme

Termdat

Slovenia:
Evroterm

UZEL, Basque country:
Euskalterm

**Termcat
Cercaterm**

Croatia:
**National Terminology Portal
(incl. Struna)**

Türk Dil Kurumu, Turkey:
Bilim ve Sanat Terimleri



"National"

- a certain **coverage** (as to contents)
- **status** (recognized by professionals and a language or terminology institution)
- **accessibility** (open and freed of copyright claims)
- large and varied **target group**

Background

"The fast development of society requires constant work on creating and making accessible agreed-upon terminologies, within more and more subject fields. An easy access to terms via the Internet in a national termbank [rikstermbank] endorses such a development."

"the establishment of a national central term bank, a "rikstermbank", is a prerequisite for an easy access to, and quality assurance of, Swedish terms in all domains."

Communications:

2005: 1 500 000 SEK; 2007: 750 000 SEK, 2009: 0;
2011: discussion about semantic resource, 2017:
development project, transfer to Language Council?

- IATE, EU; evaluation 2004
- Terminų Bankas, Lithuania & EuroTermBank

term bank, termbase ...

terminological database

term bank

database containing terminological data

[ISO 1087-1:2000]

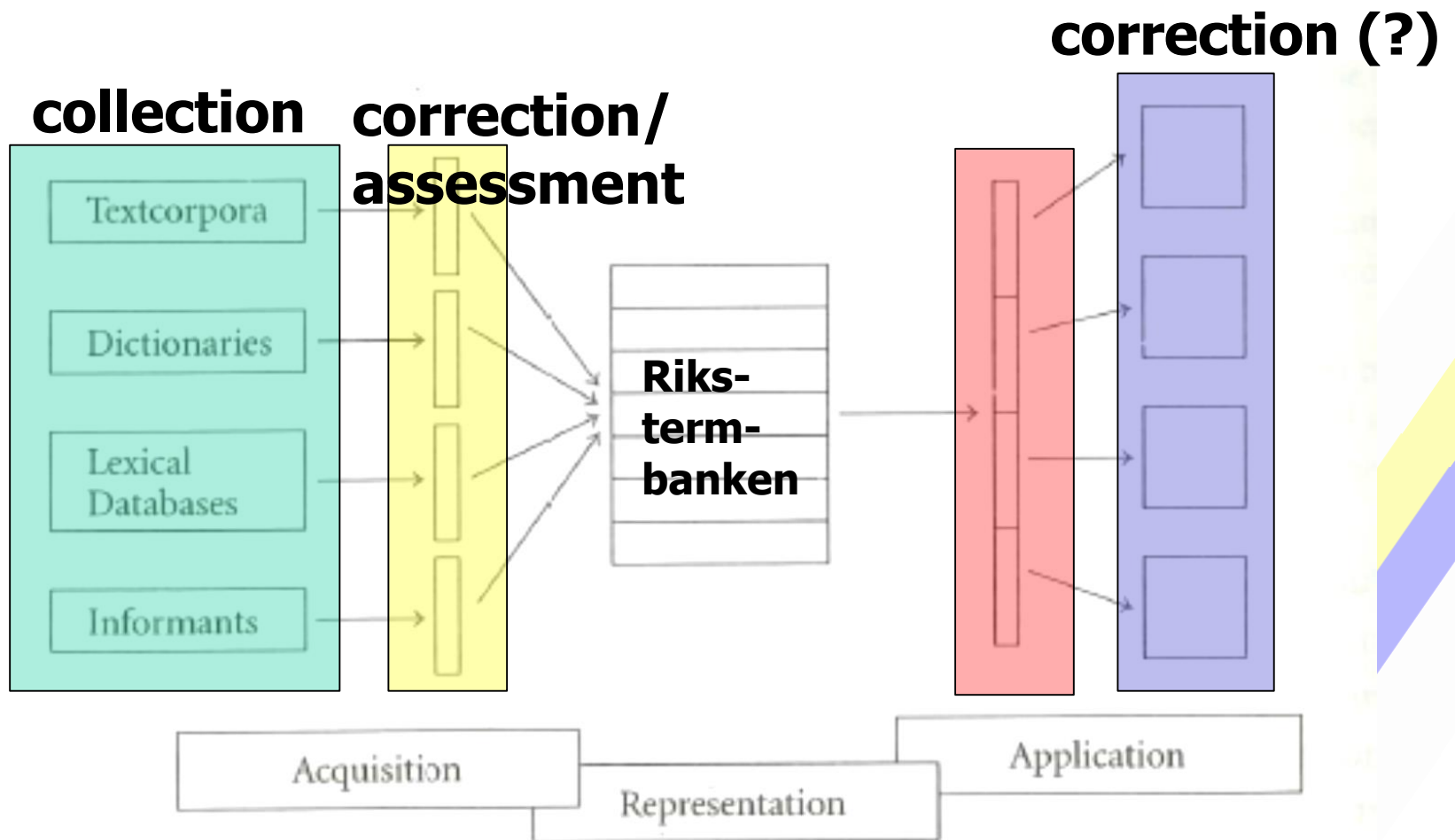
term bank

- large-scale collections of electronic term records, maintained by large institutions

term base

- smaller collection of electronic term records (and other material), maintained by e.g. a single translator (through a TMS)

[Bowker, 2015]



[Heid (1991) in Martin & van der Vliet, 2003]

Priorities

Language variety

- LSP

Languages

- Swedish or one of the official minority languages (Finnish, Yiddish, Meänkieli, Romany Chib, Sami)
- number varies according to collection

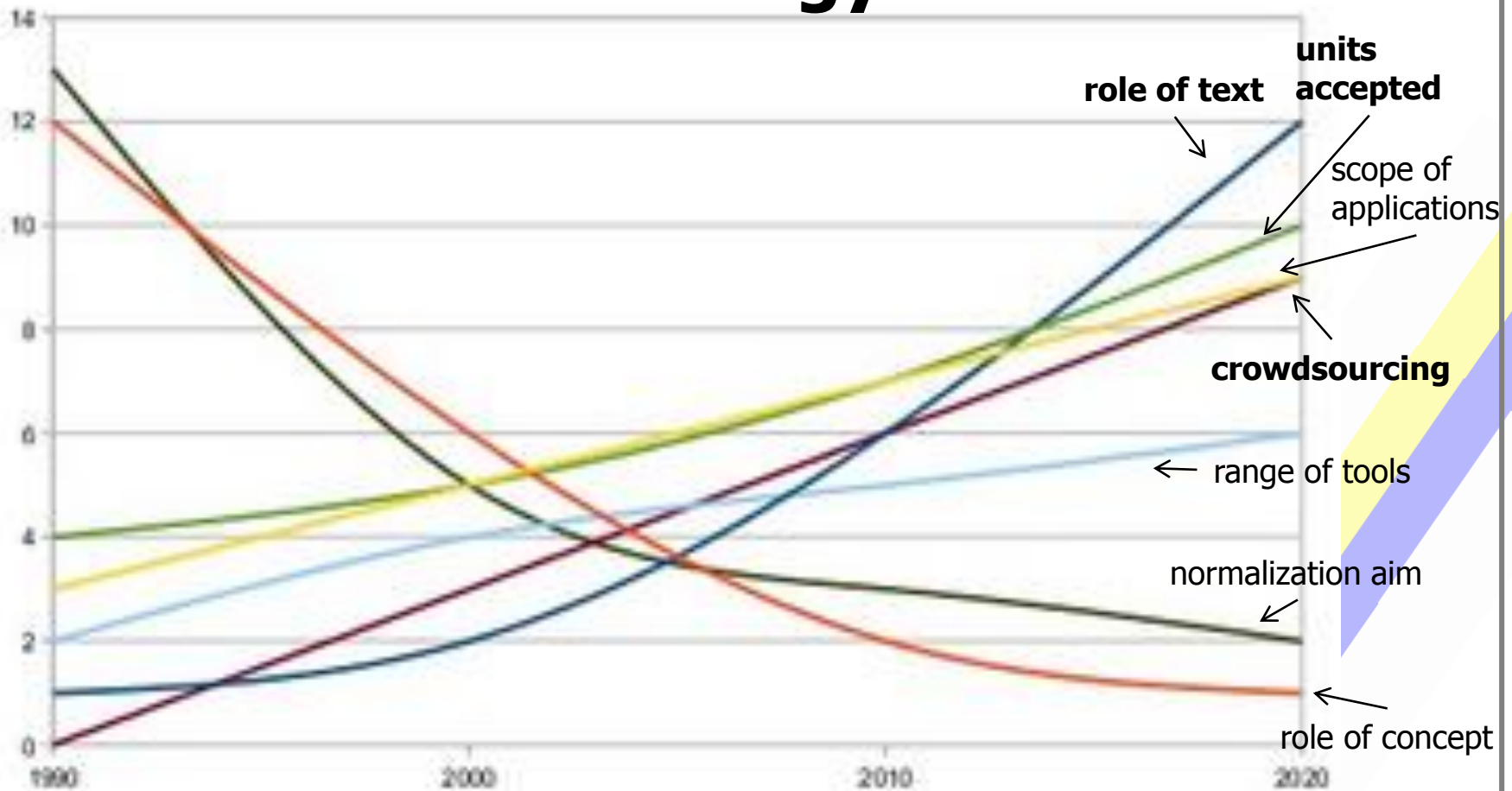
Current contents

- no limitations as to domains!
- Swedish conceptual world = starting point
- complete glossaries, but also parts of documents and excerpts
- some digitalized material
- **quality control** by terminologists (and at times the supplier)
- presentation phase → consolidation phase
overview → harmonisation
prescriptive → normative

New content: from where? how?

- commercial partners (editors, standardization bodies)
- digitalization?
- automatic extraction from text?
- change starting point: not only Swedish conceptual world ... (t.ex. material in Swedish from Finland?)
- **also** collections without Swedish?
- accept other (non-traditional) "terminological units" (TU)?
- Wikification?

Trends in terminology



[Kara Warburton, 2015]

© Terminologicentrum TNC

Trends

- Notions of “termhood” challenged
 - **term**: construct that takes shape through an analysis which gives consideration to corpus evidence, validation by subject-matter experts, and the purpose of the terminographical product
- Big data, cloud possibilities etc.
- Repurposability
- Joint efforts, communality, contributive climate (?) → crowdsourcing, nichesourcing

[Kara Warburton, 2015]

© Terminologicentrum TNC

Collection

Methodology

- weekly web search
- Google Alerts
- follow-up of statistics
- few "tips and offerings"



Using statistics ...

- **"zero hits"** = indication of what terms, domains are missing and desirable for users
- **"most sought after"** = indication of **diverse terminology needs, "trendy" terms**
- time perspective: work tool or free time entertainment?
- geography and language: multilingual interface? other languages?
- user behaviour: search strategies, "beyond the hit list"
- marketing

Import process

1. inventory (weekly) & preliminary assessment
2. formal inquiry
- 3. collection**
4. formatting
5. review, **correction/enhancements**
6. (feedback)
7. first import
8. adjustments
9. second import
10. updating

Import process (1)



FASS.se

Källan till kunskap om läkemedel

FASS.se FÖR ALLMÄNHETEN

FASS för allmänheten 

FASS för förskrivare 

FASS om djurläkemedel 

17 mars 2011

☒ Förklara svåra ord

Gå till

Min Fass 

SÖK 

Sjukdom  **SÖK**

Ordlista  **SÖK**

Läkemedelsnamn  **SÖK**

Substans  **SÖK**

Läkemedels-
grupper (ATC) >>

Miljöinformation >>

Tablett-
identifiering >>

Företag >>

LÄKEMEDELSFAKTA **LÄKARBOK** PATIENTFORUM LÄKEMEDELSUNIVERSITETET LÄNKAR INFO & HJÄLP

SJUKDOMAR LÄKEMEDEL & HÄLSA ORDLISTA

ORDLISTA

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) [Å](#) [Ä](#) [Ö](#)

Din sökning: R

Rabdomyolys

Upplösning av muskelvävnad med utfällning av myoglobin som följd vilket kan medföra akut njursvikt. Orsaken är vanligtvis mekanisk skada men kan även vara kemiskt framkallad (vissa gifter, droger och läkemedel).

Racemat

Benämning på en kemisk förening som innehåller lika delar av de molekyler som, i sin tredimensionella struktur, ser ut som varandras spegelbilder.

Racemform

Blandning av lika mängder höger- och vänstervridande isomerer av ett ämne. Kallas även racemat.

Import process (2)

TNC
TERMINOLOGICENTRUM

→ → 2009-04-06

FASS
Läkemedelsindustriföreningens Service AB
Box 17608

TNC
TERMINOLOGICENTRUM

→ → 2009-04-06

FASS
Läkemedelsindustriföreningens Service AB
Box 17608
118 92 Stockholm

FASS-ordlista-med-i-Rikstermbanken

På Terminologicentrum TNC arbetar vi nu för fullt med att förverkliga en mångårig idé: att samla all Sveriges terminologi på ett ställe, i Rikstermbanken. Rikstermbanken (med webbadress: www.rikstermbanken.se) invigdes officiellt av infrastrukturminister Åsa Torstensson den 19 mars 2009. Det är Näringsdepartementet som finansierat uppbyggnaden av Rikstermbanken och TNC som utvecklat och förvaltar Rikstermbanken. Vid öppnande innehöll Rikstermbanken drygt 250 000 termer i omkring 50 000 termposter (en termpost består vanligtvis av svensk term, definition och ibland termer på andra språk). Nu kommer termbanken gradvis att fyllas på med nytt material, och vi hoppas att många vill vara med och bidra med terminologi.

Suppliers

Approx. 400 organizations:

- authorities (majority)
- state-owned companies
- private companies
- associations
- joint terminology groups
- Terminologacentrum TNC, Språkrådet (Language Council)
- foreign organizations: Nordisk ministerråd, TSK, Nordiska språkrådet

Language Act (2009)

"§ 12 Authorities and agencies have a special responsibility for Swedish terminology within their respective domains so that such terminology is **accessible, used** and **developed**"

Import process (3)

1	Uppslagsord	Förklaring
2	abdomen	Anatomisk term för buken eller bukhålan.
3	abdominell	Som har att göra med buken eller bukhålan.
4	Abrasion	Tandslitage som kan orsakas av bl.a. tandgnissling och slipande partiklar i födan.
5	abscess	Böld, varfyllt hålrum i vävnad.
6	absolut neutrofilital	Absolut neutrofilital (ANC) är ett mått på hur väl benmärgen fungerar. ANC används bl.a. för att utvärdera olika
7	absorption	Hur läkemedlet tas upp i kroppen. □ □ Läkemedlet måste tas upp av kroppen för att kunna transporteras till det
8	Abstinens	Olika fysiska och psykiska obehagssymtom (svettningar, oro etc.) som kan uppstå vid avhållsamhet från ett lä
9	accesoriska ledningsbanor	Extra ställen i hjärtat som kan leda elektriska impulser från förmak till kammare. Normalt kan impulser bara ta
10	ACE-hämmare	Läkemedel som används för att sänka blodtrycket. ACE-hämmare minskar halten av angiotensin. Angiotensin
11	Acetylering	Adderande av en acetylgrupp till en molekyl.
12	acetylkinin	Acetylkinin är en signalsubstans och ett vävnadshormon i kroppen. Acetylkinin frisätts i nervändsluten, synaps
13	acetylkininesteras	Enzym som bryter ner signalsubstansen acetylkinin och gör det inaktivt. Acetylkinin är en förmedlare i det par
14	Acetylsalicylsyra	Acetylsalicylsyra, ASA, hämmar bildningen av prostaglandiner, vilket ger en antiinflammatorisk, smärtstillande
15	achyli	Bristande produktion av magsaft. Vanligen syftar det till bristande produktion av saltsyra i magsäcken.
16	aciditet	Surhetsgrad. Aciditet kan syfta till mängden vätejoner i en lösning vilket mäts med hjälp av pH. Ju fler vätejone
17	acidosis	Syraförgiftning. Vid acidosis sjunker blodets pH-värde under normalvärdet 7,35. Respiratorisk acidosis inträffar då
18	Acinetobacter baumannii	Bakterie som förekommer normalt hos människan. Den är ofarlig för friska människor men kan orsaka bland a
19	ackommodation	Ögats anpassningsförmåga att kunna se på olika avstånd. Ögats lins är elastisk och ändrar form efter hur myc
20	ackommodationsspasm	Kramper i ögats ciliarmuskel, det vill säga den ringformade muskeln som finns runt ögats lins, som gör att mu
21	Ackumulering	Ansamling.
22	acne vulgaris	Den vanliga formen av finnar, även kallade ungdomsfinnar.
23	acrodermatitis enteropathica	Ovanlig sjukdom som beror på oförmåga att ta upp zink i tarmen. Man får uttalade hudsymtom i form av rodn
24	Actinobacillus actinomycetemcomitans	Bakterie som kan orsaka tandlossning. Bakterien finns naturligt i munnen. Vid dålig munhygien ökar antalet b
25	Actinomyces	Bakterier som förekommer normalt i munhålan och tjocktarmen. Kan orsaka actinomycosis (kallades förr strålsv
26	acyklisk	Ej ring- eller cirkelformad. Avser oftast en molekylär struktur.
27	Addisons sjukdom	Sjukdom som drabbar binjurebarken och gör att dess funktion blir sämre. Dess normala produktion av de viktig
28	additiv effekt	Då två substanser tillsammans ger upphov till en effekt, som motsvarar summan av effekterna som substanser
29	adekvat	Tillräckligt eller vad som motsvarar förutbestämda krav i tillräcklig grad.
30	Adenin	Adenin är molekyl som ingår i DNA och RNA. I DNA sitter adenin parvis med tymin. I DNA finns vår arvsmass
31	adenohypofys	Adenohypofysen är den främre loben av en liten körtel under hjärnan som kallas hypofys. Den är bl.a. ansvarig
32	adenokarcinom	Adenokarcinom är elakartade tumörer som drabbar körtlar. Exempel på sådana tumörer är bröst- och prostata
33	adenom	En godartad tumör som drabbar körtlar. Det är ofta ganska ofarligt i sig, men kan vara ett förstadium till cance
34	adenomatoid	Körteltumörliknande.
35	adenosin	Adenosin är en kemisk förening som har flera funktioner i kroppen. Den utgör bl.a. en del av DNA där arvsmas
36	ADHD	ADHD är en förkortning för Attention Deficit Hyperactivity Disorder och är ett hyperaktivitetssyndrom med debu
37	adherent	Sammanhängande, sammanvuxen, fastklibbande. Beskriver exempelvis hur bakterier binder till cellytor.
38	adhesion	Vidhäftning eller sammanväxning.
39	adhesionsmolekyler	En grupp av molekylära strukturer, glykoproteiner, som sitter på den yttre cellväggen på till exempel en bakter
40	adipocyt	Fettcell. Den celltyp som utgör fettvävnad.
41	adjuvans	Medel som förbättrar effekten av ett annat läkemedel eller behandling. Adjuvans förekommer bland annat i vacc
42	administration	Tillförsel av ett läkemedel till kroppen. Läkemedel kan administreras många olika vägar. Peroralt (när läkemedl

Import process (4–5)

svDF varje framställning och budskap som påstår, antyder eller gör gällande att ett livsmedel har speciella näringssegenskaper vad beträffar den energi som det

- ger,
 - ger i minskad eller ökad grad, eller
 - inte ger,
- eller beträffande de näringsämnen som det
- innehåller,
 - innehåller i minskade eller ökade proportioner, eller
 - inte innehåller.

svAN Obligatorisk uppgift om ett näringsämnes kvalitet eller kvantitet utgör inte ett näringspåstående.

svTE protein

svDF ~~det~~ proteinhalt som beräknas med ledning av följande formel:

protein = sammanlagda kvävemängden enligt Kjeldahl × 6,25,

svTE kolhydrat

svDF alla kolhydrater som kan metaboliseras i människokroppen

svAN Även sockeralkoholer, sockerarter, alla monosackarider och disackarider som förekommer i livsmedel. Sockeralkoholer ingår inte i begreppet sockerarter,

svTE fett

svDF sammanlagda mängden lipider

svAN Även fosfolipider, mättade fettsyror, fettsyror utan dubbelbindning, enkelomättade fettsyror, fettsyror med en cis dubbelbindning, fleromättade fettsyror, fettsyror med metylenåtskilda cis-cis-dubbelbindningar, fiber, kolhydratpolymerer med minst tre monomerenheter som varken spjälkas eller upptas i människans tunntarm och som hör till någon av följande kategorier:

- ätliga kolhydratpolymerer som naturligt förekommer i konsumtionsfärdiga livsmedel,
- ätliga kolhydratpolymerer som framställts av livsmedelsråvaror genom fysikaliska, enzymatiska eller kemiska metoder och som har en gynnsam fysiologisk effekt som påvisats genom allmänt vedertagen vetenskaplig dokumentation,
- ätliga syntetiska kolhydratpolymerer som har en gynnsam fysiologisk effekt som påvisats genom allmänt vedertagen vetenskaplig dokumentation.

svTE genomsnittsvärde

svDF ~~det~~ värde som bäst representerar den mängd näringsämne som ett visst livsmedel innehåller och där hänsyn tas till årstidsbetingade variationer, konsumtionsmönster samt andra faktorer som kan göra att det faktiska värdet varierar

Selection:

- all or nothing – or some?
- balance?

Preparation of the contents

- termbank adaptation (reformatting according to NTRF-RTB, exclusion of remaining "book-related" aspects)
- selection
- changes for consistency
- linguistic and content-related adjustments (incl. removal of target group adaptations)
- discussion with suppliers
- illustrations
- semi-automatic three-step import control tool

Import process (4–5)

The screenshot displays a text editor window with Swedish medical text. On the right side, there is a list of terms marked as 'Borttaget' (Deleted). Red dashed lines connect specific terms in the text to their corresponding entries in the list.

Text in the editor:

- svTE: **abdomyolys**
- svDF: **upplösning** av muskelvävnad med utfällning av myoglobin som följd vilket kan medföra akut njursvikt
- svRETE: **njurinsufficiens**
- svAN: Orsaken är vanligtvis mekanisk skada men kan även vara kemiskt framkallad (vissa gifter, droger och läkemedel).
- svTE: **racemat**
- svFK: **Kemisk förening** som innehåller lika delar av de molekyler som, i sin tredimensionella struktur, ser ut som varandras spegelbilder. **dvs. har** lika mängder höger- och vänstervridande isomerer av ett ämne.
- RIAN: Jfr Biotermgruppen! / **HNI**
- svTE: **radiojodskintigrafi**
- svDF: **teknik** som möjliggör avbildning av sköldkörteln genom att injicera radioaktiva jodpartiklar som kan registreras med hjälp av en särskild kamera
- svTE: **radiojodterapi**
- svFK: **Radiojodterapi** är en typ av strålbehandling. Man äter eller dricker en liten mängd radioaktivt jod, som sedan fördelas ut i kroppen. Då jod främst används i sköldkörteln ansamlas det där och stråldosen blir därmed också högst i dess vävnad. Radiojodterapi används både för att diagnostisera och behandla olika sjukdomar i sköldkörteln.
- svTE: **strålbehandling**
- svSYTE: **radioterapi**
- svFK: **Behandlingen** görs mha. radioaktiv strålning, i syfte att döda cancerceller. Anledningen till att strålning används mot cancerceller är att de har ett sämre försvar än friska celler, och på så sätt slås ut lättare. Ofta kombineras strålning med cellgifter, och ibland också med kirurgi.

Deleted terms (Borttaget) on the right:

- Borttaget: D
- Borttaget: Benämning
- Borttaget: k
- Borttaget: ...
- Borttaget: ...
- Borttaget: R
- Borttaget: racemform
- Borttaget: B
- Borttaget: blandning
- Borttaget: ... Kallas av
- Borttaget: R
- Borttaget: En...
- Borttaget: DF
- Formaterat
- Borttaget: svSYTE
- Borttaget: DF
- Borttaget: ...
- Borttaget: Radioterapi
- Borttaget: strålbehandling
- Borttaget: ...
- Borttaget: ...
- Borttaget: ...

Import process (6)

- "livsmedel eller specifika berikningsmedel" – olika formulerat i termposterna för "berikad mat" och "berikning".
- Borde inte engelska motsvarigheten till "dietetik" vara "dietetics"?
- "sätt" verkar fattas i slutet av definitionen för "dricka" (jfr "äta"). Dessutom kan man tycka att den är väl snäv (sker det inte på kulturellt etc. sätt så dricker man inte?), men den kommer väl från ICF?
- Termposten för "enteral nutrition" är lite svårhanterlig. Handlar det om att det finns två uttryck på engelska som täcker olika mycket, men bara ett på svenska som täcker allt? Vad innebär "exempelvis" i första definitionen? Kan "enteral nutrition" användas även om oralt intag? (andra definitionen)
- ✕ TNC har ju sedan länge försökt lansera "mervärdesmat" i stället för "functional food" eftersom det är ett uttryck som är svårt att använda, böja etc. i svenskan. Används det inte alls i branschen?
- ✕ Är "ren svält" också en term?
- ✕ Skrivsättet för uttydningarna NNR och SNR varierar i texten.
- ✕ Avses inte "specifika kostråd" i termposten "kostråd"?
- ✕ Heter det "vitamin- och mineralsupplement" eller "vitamin- och mineralsupplementering"? Är det två skilda begrepp?
- ✕ Det verkar förekomma flera definitioner för "kostanamnes", bl.a. 'nutritionsanamnes med endast oralt intag' (i posten för nutritionsanamnes). Vilken är övertermen?

Import process (7–9)

Steg 1: Skicka fil inför import

Notera att filen f.n. inte

Välj fil *

-
-

Avbryt

- ⚠ Rad 0+, term=florsocker: Fält ILLT får inte kopplas till AN.
- ⚠ Rad 0+, term=kolhydratmetabolism: Fält ILLT får inte kopplas till FK.
- ⚠ Rad 0+, term=kristallisation: Fält ILLT får inte kopplas till AN.
- ⚠ Rad 0+, term=sackaros: Fält ILLT får inte kopplas till FK.
- ⚠ Rad 0+, term=sackaros: Fält ILLT får inte kopplas till FK.

Klart: Resultat efter import

St Resultat efter att data sparats.

Re

- F Namn på Kalla{optLockVersion=null, id=null, titel=Titel{titel='Ordlista om lönebildning', kortformAvTitel='Ordlista om lönebildning', overordnadTitel='null', dokumentNr='null'},
- F källa: utgivare=[Utgivare (null) Organisation: : Medlingsinstitutet Kontaktperson:: Anne-Marie Egerö Kontakt epost: : Anne-Marie.Egero@mi.se]
- F Filnamn: MI_ordlista_lönebildning_imkand.txt
- F v
- F Antal sparade 74
- F termposter:
- F
- F
- F

OK

Development in the last years

- internal consolidation (file structure, routines, etc.)
- correction of links and faults
- adding of new material
- reminders to content providers
- discussions about technological development
- project on linked open data (LOD)

Collection

Methodology

- weekly web search
- Google Alerts
- follow-up of statistics
- "few "tips and offerings"

Types

- glossaries
- excerpts from legal documents (no ©)
- material from terminology networks
- project material



Typology

1. **Copycats**

works that have been either photocopied or directly copied from a text file and then placed on an electronic platform, frequently as PDF files

2. **Faster horses**

[works] where the lexicographers do make use of the new technologies, but only in a very restricted way [IATE]

3. **Stray bullets**

[works that] use new technology, but often "in the wrong direction" ("lexicotainment") [EcoLexicon]

[Fuertes-Oliverira & Tarp, 2014]



Search



Contextual domains

No domain

Spanish



Feedback

Login

Definition

Geological formation: fundamental unit of lithostratigraphy, consisting of a certain number of rock strata that have a comparable lithology, facies or other similar properties.

Terms

- geological formation
- formación geológica
- geologische Formation
- γεωμορφή
- γεωλογικός σχηματισμός

Resources

- Formación geológica a 46 km. De Trinidad

Conceptual categories

- C.2.4 Geographic feature
- C.1.1.6 Geographic feature

Categories hierarchy

Phraseology

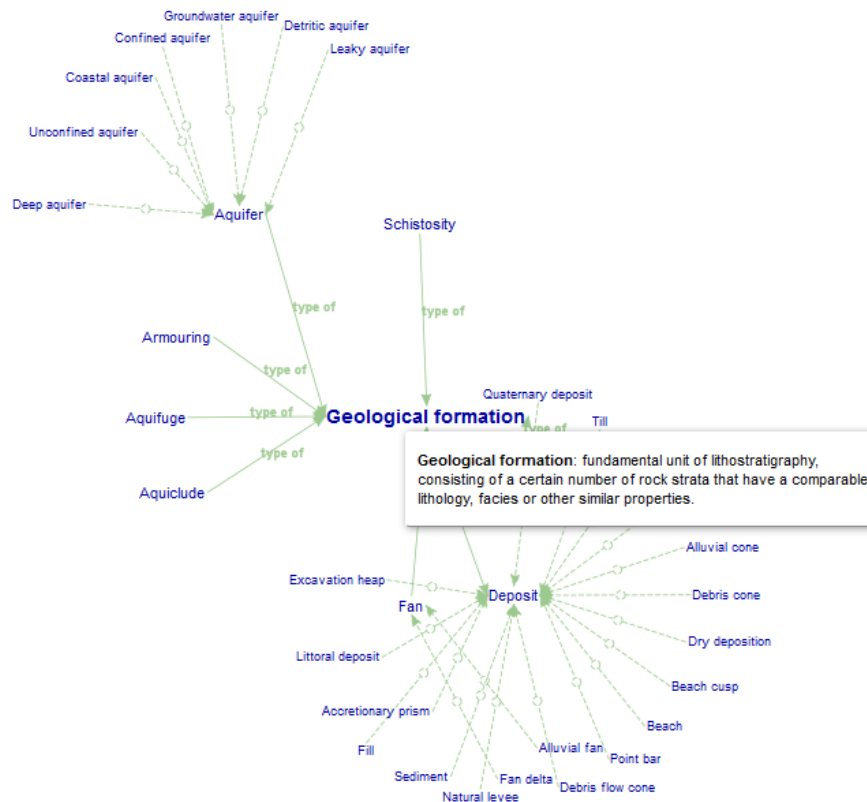
History

Search results

A-Z

Path

Search concordances



Typology

1. Copycats

works that have been either photocopied or directly copied from a text file and then placed on an electronic platform, frequently as PDF files

2. Faster horses

[works] where the lexicographers do make use of the new technologies, but only in a very restricted way [IATE]

3. Stray bullets

[works that] use new technology, but often "in the wrong direction" ("lexicotainment") [EcoLexicon]

4. Model T Fords

works [which] have gone beyond the traditional boundaries and have not only made use of the existing technology in order to provide quicker data access, but also to adapt the dictionary articles to the various functions displayed by the dictionary

5. Rolls Royces

e-tools which permit individualized solutions for specific and individual users in concrete situations, and which might also combine access to selected data in a prepared database

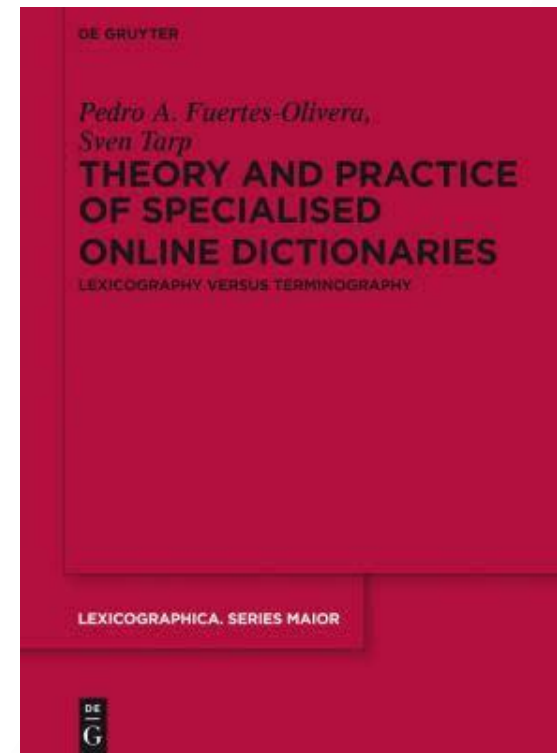
[Fuertes-Oliverira & Tarp, 2014]

Assessment/evaluation

assessment

<terminology> process to demonstrate that a terminological resource fulfils specified requirements [ISO 23185]

- Authors' views
- Functionalities
- Access modes
- Internet technology
- Lexicography
- Production cost
- Information cost
- Updating
- Expert participation
- Data selection



[Fuertes-Oliveira & Tarp, 2014]

Assessment: steps

- Specifying the goals
- Identifying benchmarks
- Checking conformity to benchmarks
- Test-case assessment

[ISO 23185]

© Terminologicentrum TNC

Type of content

terminological resource

terminological data resource

the entity composed of **collections** of terminological data associated with its inherent usability attributes that make the terminological data potentially useful and suitable to the users' need

usability attribute

a property related to the usability of terminological data

[ISO 23185]

Evaluation of sources

“From the point of view of a user, the following are the two main requirements for usable terminological data:

- 1) the terminological data **meets the user's need to acquire terminological information content**;
- 2) the terminological data are **designed to allow desired access or processing**.

Terminological resources shall be assessed on the basis of analyses of their **creation**, their **management** and their **potential use**. [...]

[ISO 23185]

© Terminologicentrum TNC

Evaluation of sources (cont.)

“When a terminological resource is being used, the **content pertinence, structural conformity and formalistic quality** of the terminological data are the greatest concerns for users. [...] The word “quality” here refers to the **content quality** of terminological data that a terminological resource holds. Two aspects are concerned: its level of **quality from the standpoint of linguistics/language** and its level of **quality from the standpoint of specialized knowledge**.”

[ISO 23185]

© Terminologicentrum TNC

Evaluation of sources

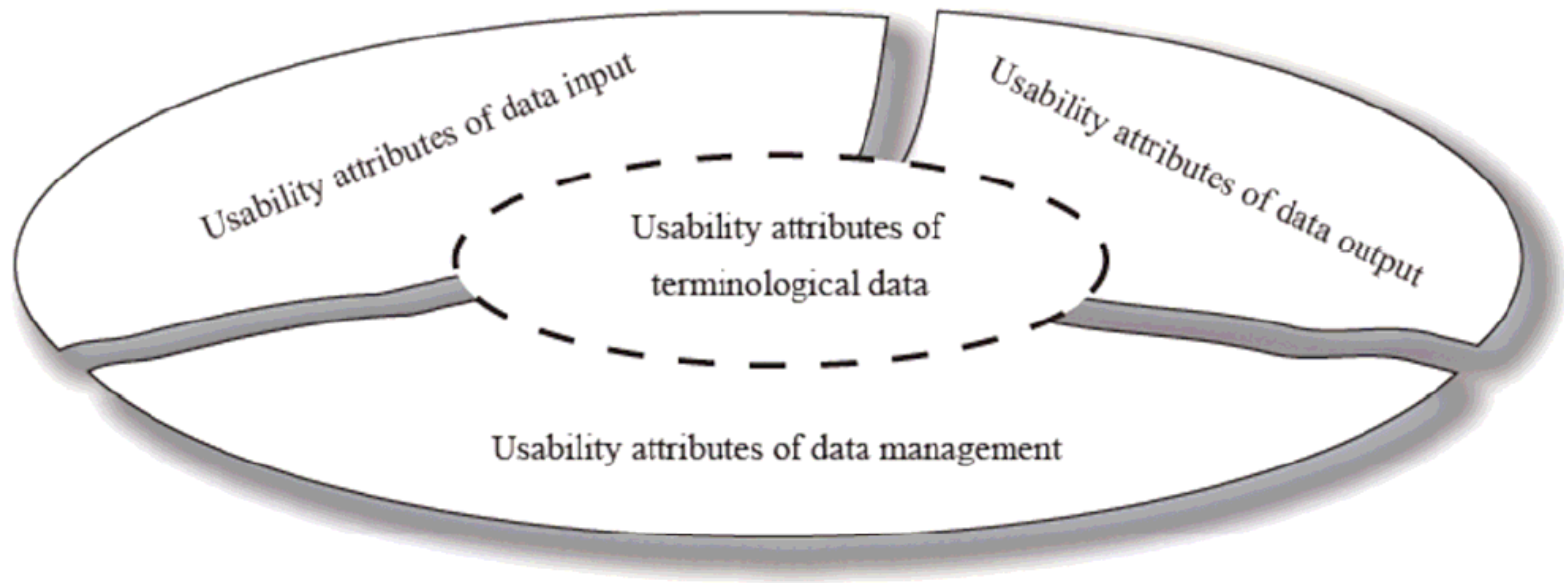


Figure 1 — Four sets of usability attributes

[ISO 23185]

© Terminologocentrum TNC

Usability attribute (1)

related to terminological data

A terminological resource shall have enough attributes to meet the complexity of the terminological data while at the same time meeting the requirements for specific purposes, such as

- data structure specification
- data category coverage
- subject-field coverage
- language coverage
- compliance with rules of coherence
- **use of controlled external data**
- **use of authoritative sources**
- intellectual property rights ownership indication
- symmetry of the terminological data collection
- size of the terminological data collection

[ISO 23185]

© Terminologicentrum TNC

Collection!?

Pros

- control
- consistency ...
- current quality not (terminologically) acceptable
- not structured or accessible formats

Cons

- labour-intensive
- double storage → varying contents for same source



State-of-the-art

(Swedish) online LSP glossaries

- present an authentic, up-to-date language use, not available anywhere else
- are often developed by non-lexicographers for non-lexicographers
- are typologically diverse and designed by purpose rather than by principle
- are translation- or fact-oriented
- rarely contain any explicit linguistic information (inflection, gender, pronunciation, etc.)

Online LSP glossaries work well in spite of the formal shortcomings. It should therefore be possible to enhance them, both quantitatively and qualitatively, through simple means (e.g. links to good examples, basic manuals).

[Törnqvist, 2007]

© Terminologocentrum TNC

Rikstermbanken's initial criteria ...

- origin
 - contents
 - structure and form
 - format
 - uniqueness
 - timeliness
-
- assessment by two terminologists

Correction



Common flaws = challenges

- poor contents
- unstructured data
- spelling mistakes
- "print" character
 - alphabetical ordering
 - "Capital Mistake"
- "hanging" references, unclear intra-crossreferences
- synonym conflicts, variants
- dissimilarity of related concepts



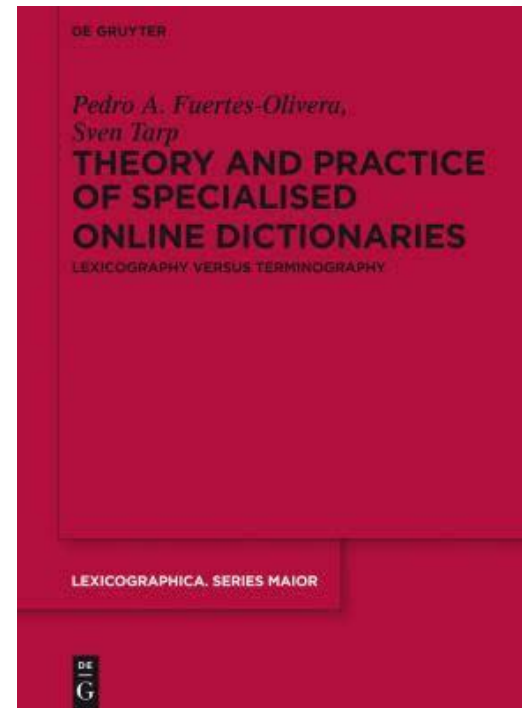
Entry layout

SVENSK TERM (REKOMMENDERAD):	uppkolning
SVENSK TERM (AVRÄDD):	inkolning
DEFINITION:	höjning av kolhalt
ANMÄRKNING:	Inom värmebehandlingstekniken vanligen indiffusion av kol i yta som led i sätthärdning. Uppkolning tillgår vanligen så, att materialet hålls vid en temperatur över A_{c1} i närvaro av ett kolhaltigt medel i form av gas, vätska eller pulver.
SE ÄVEN:	enkelhärdning, diffusion, direkthärdning, dubbelhärdning
ENGELSK TERM:	carburizing carbon cementation
DEFINITION:	raising of carbon content
ANMÄRKNING:	In heat-treating practice usually the diffusion of carbon into the surface as part of case hardening. In the carburizing process, the material is held at a temperature above A_{c1} in the presence of a carbon-rich gas, liquid or powder.
TYSK TERM:	Aufkohlen (n)
FRANSK TERM:	carburation (f) cémentation par le carbone (f)
SPANSK TERM:	cementación
DANSK TERM:	opkulning
RYSK TERM:	науглероживание
JAPANSK TERM:	sintan
KÄLLA:	Terminologikum TNC: Värmebehandlingsordlista 1974

User reaction: "It seems to be very user friendly. The system is fast, and the records uncluttered and easy to read."

Assessment/evaluation ...

"To the best of our knowledge [...] the translation of terminological theories into real and working terminographical products, has so far a lot left to be desired."



[Tarp & Fuertes-Oliveira, 2014]

© Terminologicorum TNC

Difficulties of a terminologist

"The terminologists must consider himself a reporter, not an inventor. If he were an inventor, he might set up a terminology all his own, consisting preferably of terms not in everyday use and not employed in any other field. [...]"

The compiler of a terminology in a special field must, like the compiler of a general dictionary, describe what he finds."

[Tenner, **1941**]

© Terminologicentrum TNC

Differing starting points

TERMINOLOGY

onomasiological starting point

starts from the contents (the concept) and moves towards the linguistic expression (the term)

term



concept

LEXICOGRAPHY

semasiological starting point

starts from the linguistic expression and determines its meaning

Term bank contents: challenges

- Selection: all or nothing – or a little?
- Interpretation of contents
- Term choice (variants, synonyms etc.)
- Definition vs. explanation
- Updating vs archiving → consistency changes?
- “Record breaking” & “record making”
- Incorporated (look-up) terms
- Document types: legal documents ...
- Updating

"Record breaking"

Before

incidens	1. antalet fall av en viss sjukdom som uppträder i en befolkning under viss tid; anges t ex som antalet diagnoser per 1 000 invånare per år. 2. antalet av en viss studerad »händelse i en klinisk prövning eller kohortundersökning, dividerat med antalet deltagare i gruppen. Graden av skillnad mellan två grupper incidensstal kan uttryckas genom att det ena divideras med det andra till en incidenskvot.
----------	---

After

SVENSK TERM:	incidens (1)
FÖRKLARING:	Antalet fall av en viss sjukdom som uppträder i en befolkning under viss tid; anges t.ex. som antalet diagnoser per 1 000 invånare per år.
SE ÄVEN:	risk
KÄLLA:	SBU, Statens beredning för medicinsk utvärdering, Jörgen Malmquist: Ordlista 2004 Synpunkter på denna termpost?
SVENSK TERM:	incidens (2)
FÖRKLARING:	Antalet av en viss studerad händelse i en klinisk prövning eller kohortundersökning, dividerat med antalet deltagare i gruppen. Graden av skillnad mellan två grupper incidensstal kan uttryckas genom att det ena divideras med det andra till en incidenskvot.
SE ÄVEN:	händelse
KÄLLA:	SBU, Statens beredning för medicinsk utvärdering, Jörgen Malmquist: Ordlista 2004 Synpunkter på denna termpost?

viss tid; anges

ortundersökning,
grupper
en incidenskvot.

"Record making"

Regeringskansliets ordlista

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Å Ä Ö

A

Allmän beredning

Diskussion mellan statsråden i regeringen i omfattande eller principiellt viktiga frågor. Allmän beredning hålls normalt på torsdagar efter regeringssammanträdet.

- Veckans regeringssammanträde
- Handläggning och beredning av regeringsärende

Allmän handling

En handling är allmän om den har kommit in till eller upprättats vid en myndighet och förvaras där. Allmänna handlingar ska enligt offentlighetsprincipen vara offentliga och tillgängliga för alla att läsa, om inte sekretess föreligger. Se även handling, offentlighetsprincipen och sekretess.

- Offentlighetsprincipen
- E-post till regeringen och departementen

Anslag

Belopp i statsbudgeten som riksdagen beslutar får användas för ett visst ändamål.

^ Upp

^ Upp

"Record making" (cont.)

Kleines Wörterbuch der Kanzlei der Ministerien

A B C D E F G H I J K L M N O P Q R S T U V W V W X Y Z Å Ä Ö

A

Allgemeine Beratung

(allmän beredning)

Diskussion der Minister in der Regierung in umfangreichen oder prinzipiell wichtigen Fragen. Eine allgemeine Beratung findet normalerweise donnerstags nach der Kabinettsitzung statt.

Allgemeines Dokument

(allmän handling)

Siehe offizielle Akte.

Allgemeine Wahlen

(allmänna val)

Die Wahlen zum Reichstag, zum Gemeinderat und zum Provinziallandtag werden alle vier Jahre am dritten Sonntag im September abgehalten. Die Wahl zum Europaparlament findet alle fünf Jahre im Frühjahr statt. Zwischen den ordentlichen Wahlen können zusätzliche Wahlen zum Reichstag abgehalten werden.

Bewilligung von Mitteln

(anslag)

Ein Betrag im Staatshaushalt, der auf Vorschlag des Reichstags für einen besonderen Zweck zur Anwendung gelangen darf.

^ Upp

Record making (cont.)

SVENSK TERM:	anslag
DEFINITION:	belopp i statsbudgeten som riksdagen beslutar får användas för ett visst ändamål
ENGELSK TERM:	appropriation
DEFINITION:	a sum of money in the central Government budget which the Riksdag decides may be used for a specific purpose
TYSK TERM:	Bewilligung von Mitteln
DEFINITION:	ein Betrag im Staatshaushalt, der auf Vorschlag des Reichstags für einen besonderen Zweck zur Anwendung gelangen darf
FRANSK TERM:	crédits
DEFINITION:	montant budgétaire que le Riksdag a décidé d'utiliser pour un objectif précis
BOSN./KROAT./SERB. TERM:	izdvajanje iz proračunskih sredstava
DEFINITION:	dio novca iz državnog proračuna za koji Riksdag odluči da se može iskoristiti za određenu svrhu
POLSK TERM:	przyznane środki
DEFINITION:	kwota w budżecie państwa, która decyzją parlamentu może zostać wykorzystana na określony cel
FINSK TERM:	määräraha
DEFINITION:	rahasumma valtion budjetissa, jonka valtiopäivät on päättänyt myöntää johonkin tarkoitukseen
KÄLLA:	Regeringskansliet: Regeringskansliets ordlista 2005–2007

Updating

- various levels: spelling mistake → new definition
- difficult to keep track of without reporting system; automatic system?
- keep older versions/archived?
- "ontologization", i.e. enhance what is there?
- "negative" updating?

Terminology and resource harmonisation – TNC's experiences

- on a national level: **Rikstermbanken**
- background & perspectives & user survey
 - content revision
 - harmonisation **within a source**
 - definition – explanation
 - harmonisation **between sources** (i.e. within the termbank as a whole)
 - doublettes
 - problems and solutions
 - content presentation
 - content updating
- user adaptation in term banks

Challenge: harmonizing content

- within and between sources
- signalize various statuses ("primaries")
- automatic handling of doublettes
 - automatic calculation of "definition similarity"?
- version management
- automatic updating of content
- automatic notification of updating (to users, of existing links etc.)

Harmonisation: within a source

- often semasiological presentation → redundancy (e.g. synonyms in separate records)
- choice of definition or explanation
- with respect to macrostructure (cross-references etc.)
- homonyms

Harmonisation: between sources

- (automatic) removal of absolute doublettes (but other information, other languages etc.?) →
- limit (%) of "definition similarity" – calculation?
- combination of several **sources** in one record instead?
 - several organizations using the same definition is in itself an interesting piece of information → special marking in hit list?
- source respect? © issues?

Legitimate redundancy?

Resultat: 64 träffar vid sökning efter "absorption".

absorption SV

FÖRKLARING: En process där t.ex. en foton fångas upp av ett material och försvinner. ...

KÄLLA: Lunds tekniska högskola, Lunds universitet: Begreppslista | 2008, 8. Begreppslista, s. 173–182

[Visa hela termposten](#)

absorption SV

FÖRKLARING: En ljus yta reflekterar det mesta ljus som träffar den, medan en mörk yta ...

KÄLLA: Institutet för infologi, Rune Pettersson: Informationsdesign från a till ö | 2008

[Visa hela termposten](#)

absorption SV

FÖRKLARING: Hur läkemedlet tas upp i kroppen. Läkemedlet måste tas upp av kroppen för ...

KÄLLA: Fass-verksamheten, LIF: Ordlista | 2008

[Visa hela termposten](#)

absorption SV

DEFINITION: uppsugning eller upptagning

KÄLLA: Formas: Ord- och begreppsförklaringar | 2004, s. 100–104

[Visa hela termposten](#)

absorption SV

DEFINITION: den process som innebär att ljus och annan strålning omvandlas till värme ...

KÄLLA: Energimyndigheten: Ordlista | 2003

[Visa hela termposten](#)

absorption SV

DEFINITION: process som åstadkommer energiförlust hos strålning när den passerar ...

KÄLLA: MSB; Myndigheten för samhällsskydd och beredskap: Ordlista | 2003, s. 78–97

[Visa hela termposten](#)

absorption SV

DEFINITION: ett materials upptagande/omvandlande av strålningsenergi till värmeenergi

KÄLLA: Formas: Ordlista | 1995, s. 213–227

[Visa hela termposten](#)

- varying characteristics, varying sources
- national termbank: another purpose
- lack of classification complicates matters (cf IATE and EuroTermBank)
- solution:
 - leave untouched – to show variation
 - combine (some) – but how?
 - grading system (cf IATE)?

Correction!?

Pros

- = enhancement!
- consistency
- gateway to new contacts
- current quality not (terminologically) acceptable



Cons

- labour-intensive
- risk of changing contents
- varying contents for the same source

termbank = user-created product?

- contents (comments, etc.)
- technology
- presentation (interface, etc.)

participatory design

process that involves developers, business representatives, and users working together to design a solution

It actively involves users in the design process to help ensure that the product designed meets their needs and is usable in the process.

[<http://www.usabilitybok.org/glossaryF>]

Customization choices

- Search (preferences, profile):
 - very common
- Hit list:
 - common
- Record display:
 - macro- and microstructure: varies
 - fixed (vertical) order: common
 - user can choose display of various fields: Termium, ISO CDB
 - "metalanguage": almost always
 - field granulation: sometimes (More ...)
 - graphic means: quite common (but WAI etc.)
- RTB: none! = old-fashioned?

”Placing greater trust in users than earlier seems to be a widely accepted point of view”

[Tavast, 2008]

”Allowing different views for different users is important in order to win over your users.”

[ISO/CD 26162:2009]

”Actually, the composition of the article is not at all relevant to the answer to many questions a user can ask a dictionary.”

[Oppentocht & Schutz, 2003]

© Terminologicentrum TNC

Adaptation to user needs!?

- = Important for terminology products!
- But: sometimes over-estimated, esp. concerning human users and layout of term banks?
- Demand, frequency of usage vs development costs?

crowdsourcing

method of outsourcing a function originally undertaken by employees to a group ("crowd") of external contributors either as an open call or in a more restricted way

[Baines, 2013]

collaborating with an indeterminate group to carry out development in research

[Hakkarainen]

the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers

[Merriam-Webster]

© Terminologikum TNC

nichesourcing

crowdsourcing where tasks are distributed amongst a small crowd of citizen scientists (communities)

[Jussi-Pekka Hakkarainen,
http://blogs.helsinki.fi/natlibfi-bulletin/?page_id=206]

© Terminologikum TNC

Contents

"traditional"

- terminologist = content provider & content checker
→ consistency, concept-orientation guaranteed
- ready-made sources presented together; result of systematic work (?)
- divided into more data categories (TBX etc.)

wiki-based

- terminologist = moderator
- content = created continuously + ready-made; result of ad hoc work (?)
- many (?) involved in content creation → fragmentation risk (?)

Presentation

"traditional"

- user adaptation (often) necessary
- more modifiable
- more advanced options
- graphic means to represent concept systems
- term record = basis

wiki-based

- no user adaptation necessary (easily recognizable) (?)
- less fine-grained
- easily multilingual, but independent fields/language

Rikstermbanken and open data?

Need: access through other means than web interfaces.

In what way? Linked open data (LOD)?

Today's web

Google "*James Joyce*"

Is it a person, a work of art or a city? Google can guess, but doesn't know.

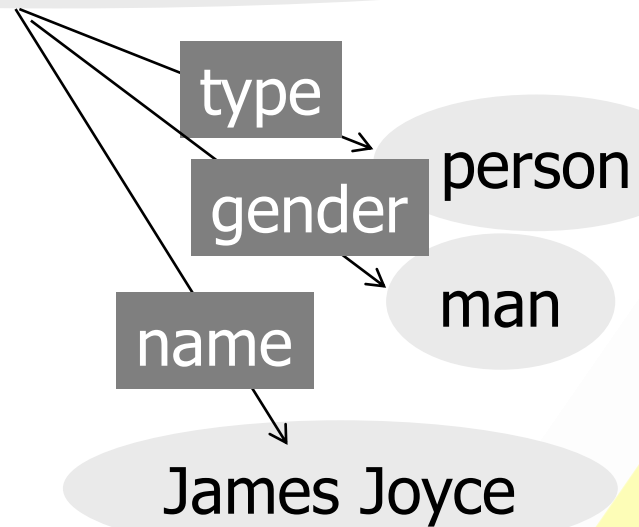
Semantic web

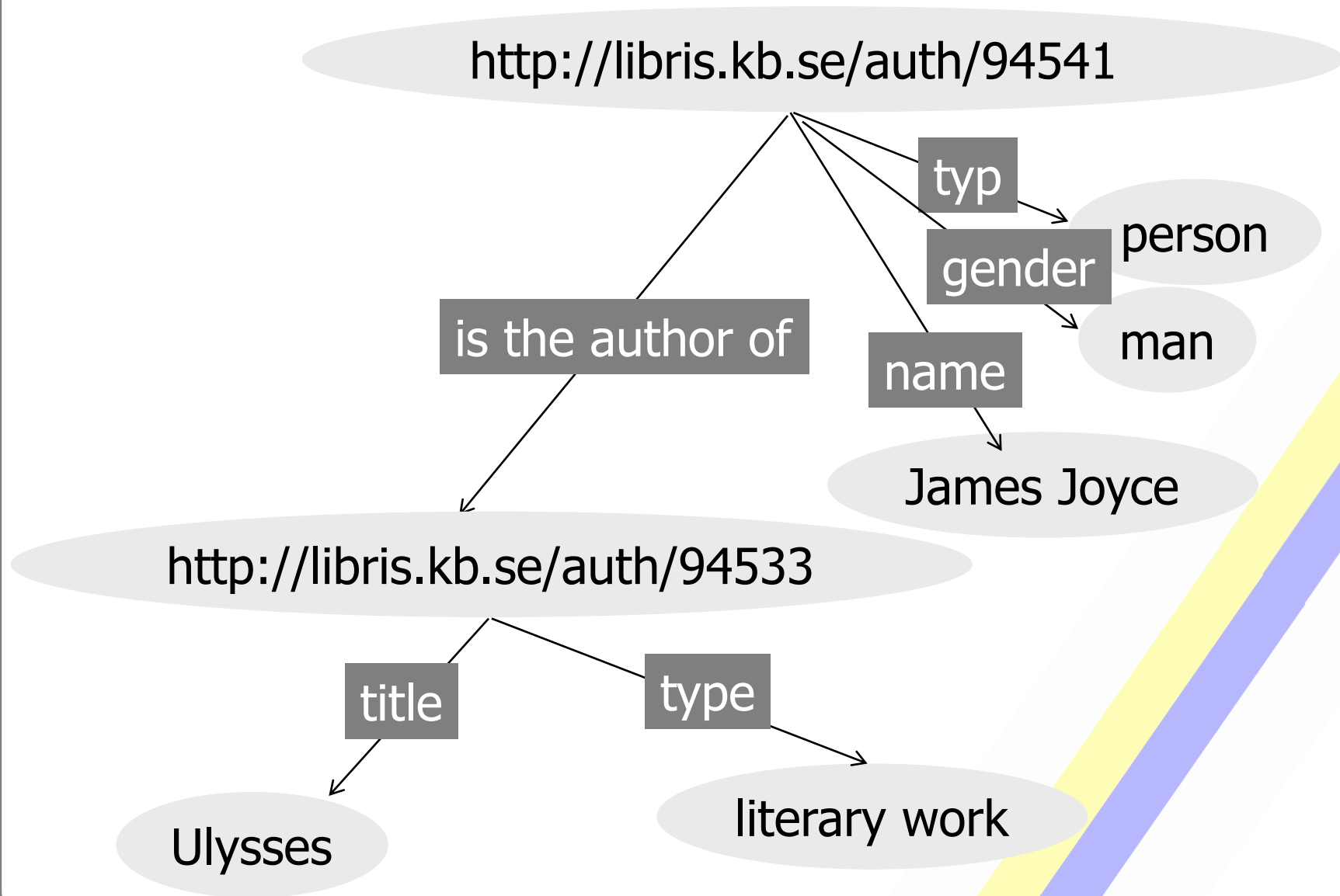
James Joyce is considered a *resource* with a unique identifier (= a unique URL), e.g.
<http://libris.kb.se/auth/94541>

To this are attached a number of facts:

- that it is a person
- the gender of the person
- the full name
- osv.

<http://libris.kb.se/auth/94541>





Linked open data (LOD)

linked open data

(on the web:) linked data which are also open data

linked data

(on the web:) structured data which follow RDF-rules and which are made available through the HTTP-protocol

open data

data which are publicly available via internet and which can be used, re-used and distributed freely and without cost

[Öppna företagsdata – Skatteverket 2012]

What's new?

Not a closed data base with everything about Joyce

Instead: resources and links on the net

New way of thinking about, and managing, data

<http://lod-cloud.net/>

Different worlds

World of semantic web

Vocabularies, ontologies

Terms, hierarchies, relations, seldom definitions

World of terminology:

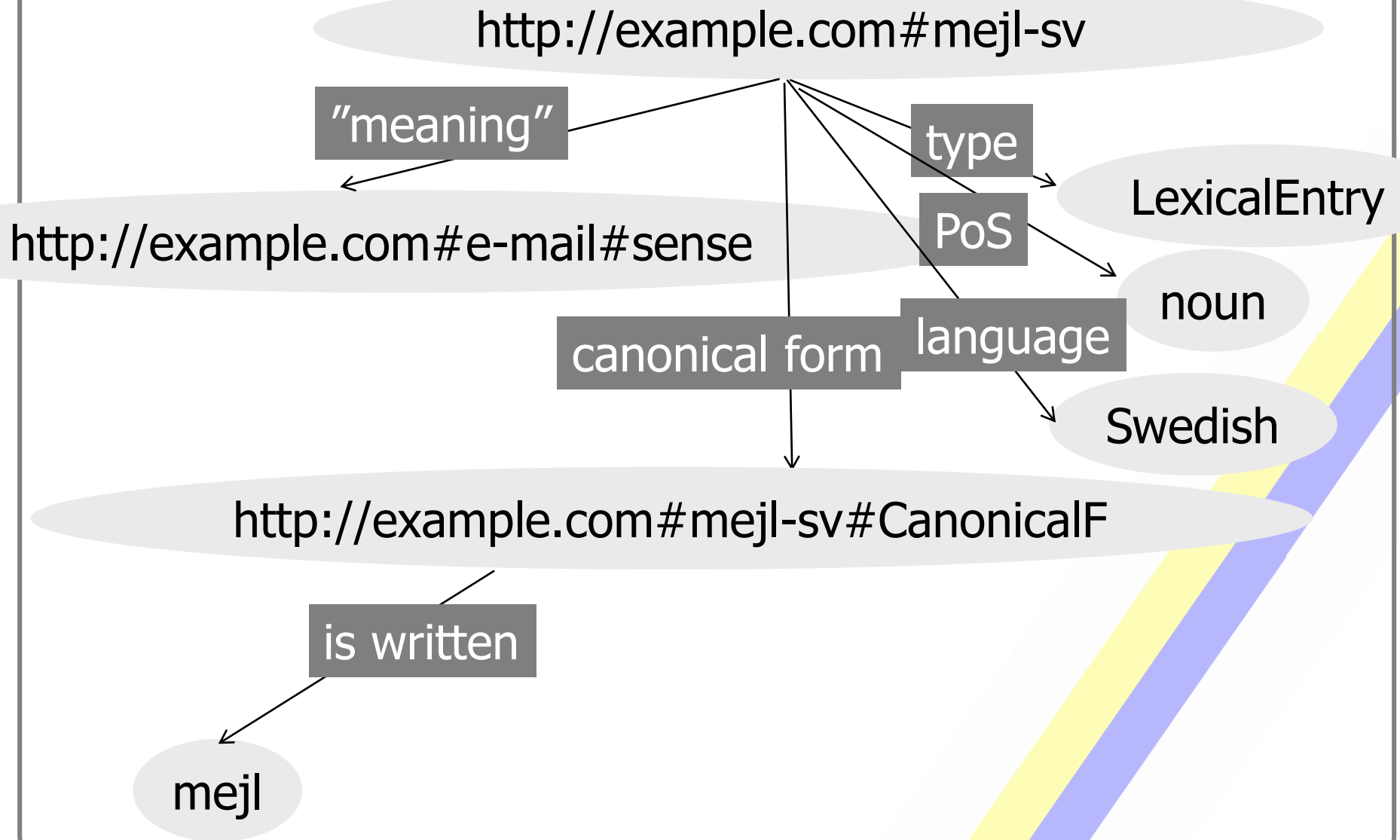
Concepts, concept systems, terms

Often definitioner.

These worlds rarely meet, how can this be facilitated?

Terminology as LOD?

- Terminology is not lists with names, addresses, geographical data, etc.
- Terminology needs standardizing – and who should be doing that? Acceptance?
- What is the best technology?





TNC

Västra vägen 7 B, Terrassen
169 61 Solna

Tfn: 08-446 66 00

Fax: 08-446 66 29

www.tnc.se

henrik.nilsson@tnc.se



RIKSTERMBANKEN

www.rikstermbanken.se

**Rikstermbankssekretariatet:
rikstermbanken@tnc.se
+46 8 446 66 00**