



## WordTies - a web interface for browsing wordnets across languages

Bolette Sandford Pedersen  
University of Copenhagen



## What are wordnets?

**WordNets** are lexical databases compiled for different languages, typically inspired by [Princeton WordNet](http://en.wikipedia.org/wiki/Princeton_WordNet).

Wordnets group words into sets of synonyms, provide short, general definitions, and record the various semantic relations between these synonym sets.

The purpose of wordnets is twofold: to produce a combination of dictionary and thesaurus of a particular language that is more intuitively usable, and to support language technology applications that include some degree of text understanding

(Based on <http://en.wikipedia.org/wiki/WordNet>)



## What is WordTies?

WordTies is a web interface developed to visualize monolingual wordnets as well as their alignments with wordnets in other languages.

In the WordTies browser, semantic relations are made available in an intuitive and graphical way.

WordTies was developed as a prototype in the META-NORD project (no. 270899) with the Nordic and Baltic countries. Technical developers were Anders Johannsen and Mitchell Seaton.



## What does WordTies include?

WordTies currently includes Danish, Finnish, Estonian and Swedish wordnets. Polish wordnet is being included.

For details on WordTies, see

<http://www.ep.liu.se/ecp/085/016/ecp1385016.pdf>

published in *Proceedings from the 19th Nordic Conference on Computational Linguistics (NODALIDA)*. Linköping Electronic Conference Proceedings; Volume 85.



## How to use WordTies

- Go to the WordTies website: [wordties.cst.dk](http://wordties.cst.dk)
- Read the brief intro and decide which wordnet to browse as your source
- You can now choose between browsing the full monolingual wordnet or only the concepts that are aligned with other wordnets
- A drop-down list ensures that you are informed about aligned concepts
- Further, you can either browse the wordnet via searches in the language of the chosen wordnet or you can search the wordnet via English
- The selected concepts can be browsed in different ways and you can move around in the semantic hierarchy by clicking on each concept
- If the wordnet includes ontological types as features, these are presented at the bottom of the page
- If you have chosen to view alignments with other wordnets, these alignments will become visible and you can click into the aligned concepts and view in more detail how these are described in the target language



# Example search in WordTies

Go to the WordTies website and select DanNet:

## WordTies: A Nordic/Baltic Multilingual Wordnet Initiative

**WordTies** describes a multilingual wordnet initiative embarked in the [META-NORD/META-NET](#) projects and concerned with the validation and pilot linking between Nordic and Baltic wordnets.

### Wordnets in Nordic/Baltic countries

The builders of these wordnets have applied very different compilation strategies: The Danish, Icelandic and Swedish wordnets are being developed via monolingual dictionaries and corpora and subsequently linked to Princeton WordNet. In contrast, the Finnish and Norwegian wordnets are applying the expand method by translating from [Princeton WordNet](#) and the Danish wordnet, DanNet, respectively. The Estonian wordnet was built as part of the EuroWordNet project and by translating the base concepts from English as a first basis for monolingual extension.

**CLARIN ERIC**  
Common Language Resources and Technology Infrastructure



WordTies has been included as a [CLARIN ERIC](#) showcase.

### Aim of multilingual wordnets

The aim of the multilingual action is to test the perspective of a multilingual linking of the Nordic and Baltic wordnets and via this (pilot) linking to perform a tentative comparison and validation of the wordnets along the measures of taxonomical structure, coverage, granularity and completeness. WordTies currently includes [Danish](#), [Finnish](#), [Swedish](#) and [Estonian](#) wordnets which have been linked to [Princeton Core WordNet](#), thereby providing a common, linked coverage of all in all 5,000 core synsets. In the current web interface, 20% of these have been manually validated and are made visible through multilingual links.

Select an available Nordic/Baltic *source* wordnet to browse below:

- [DanNet](#) (Danish Wordnet)
- [FinnWordNet](#) (Finnish Wordnet)
- [TEKsaurus](#) (Estonian Wordnet)
- Swesaurus\* (Swedish Wordnet)

\*Alignments or relational links via Princeton WordNet are presented in the other



## Example search in WordTies

Search in DanNet via English and select the concept *bread*:

wordties

☐ Search in full DanNet
 ☐ CorePWN alignments  
☒ English search
 ☐ Multilingual alignments

### About DanNet



The conceptual network displayed in this website is based on data from [DanNet](#) version 2.2. The project is a collaboration between [Center for Sprogteknologi](#) at University of Copenhagen and [Det Danske Sprog- og Litteraturselskab](#).

In total there are registered 66,308 concepts which is held in place by 323,586 relations, loaded in *WordTies*.



DanNet's is influenced by the English language [WordNet](#) since the mid-1980s has been developed at Princeton University.

### About WordTies

The *WordTies* application is most useful as a utility software for browsing translational links between the different wordnets. Nevertheless, an additional ambition is to facilitate better browsing into the particular wordnets and thereby enable evaluation of differences in structure and granularity.

[WordTies](#) has been included as a [CLARIN ERIC](#) showcase.  
 Developed by [Anders Johannsen](#) and [Mitchell Seaton](#).  
 This site presents data from [DanNet](#).

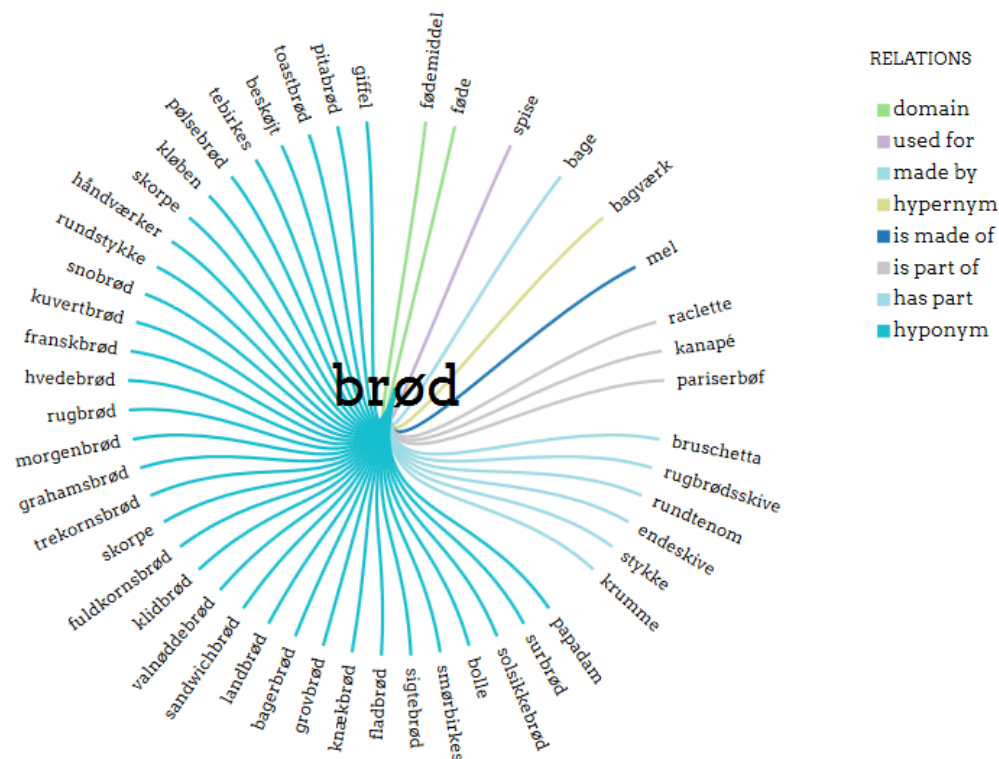


# Example search in WordTies

Browse the concept “brød”, (*bread*)

## brød 4

fødemiddel fremstillet ved at en dej af mel, væske ...





## Example search in WordTies

Consider in more detail the sub-concepts of *bread* and their definitions, for instance the typical Danish bread “rugbrød” (*ryebread*) as well as the popular Indian bread type “papadam”:

- kuvertbrød: lille brød der spises til en forret, fx suppe el. ...
- landbrød: stort (rundt) brød der er bagt med sigtemel el. an ...
- morgenbrød: bagerbrød som man spiser til morgenmad
- papadam: papirtyndt, sprødt indisk brød på størrelse med en ...
- pitabrød: lille fladt, rundt og hult brød der som regel skær ...
- pølsebrød: lille blødt, aflangt brød som serveres til kogte e ...
- rugbrød and rugbrøds-: mørkt og fast (surdejs)brød med et stort indhold a ...
- rundstykke: rundt stykke morgenbrød med sprød skorpe, ofte dry ...
- sandwichbrød: let, hvidt brød der sælges i kvadratiske skiver og ...
- sigtebrød: lyst brød bagt med sigtemel og hvedemel
- skorpe: lille, hårdt brød der er bagt, flækket og ristet s ...
- skorpe: et stykke tørt, gammelt brød
- smørbirkes: et stykke morgenbrød hvor gærdejen er udrullet med ...
- snobrød: lille brød af franskbrødsdej der er viklet om en a ...
- solsikkebrød: brød med indhold af solsikkekerner
- surbrød: brød der bages med grove melsorter, fx sigtemel, o ...
- birkes and tebirkes: et stykke morgenbrød af butterdej der er foldet og ...
- toastbrød: let. hvidt brød der sælges i kvadratiske skiver oa ...



## Example search in WordTies

Consider the alignments of *bread* with other languages and select the Finnish wordnet in order to compare how *bread* is described in this language:

### Alignment

The English Wordnet has concepts that correspond to the Danish:

→ (English) bread, breadstuff, and staff of life: *food made from dough of flour or meal and usually raised with yeast or baking powder and then baked*

→ (Estonian) leib : (leib) hrl. juuretise abil hapendatud ning kergitatud (rukkijahu)taignast küpsetatud toit; leivapäts; leivakäär

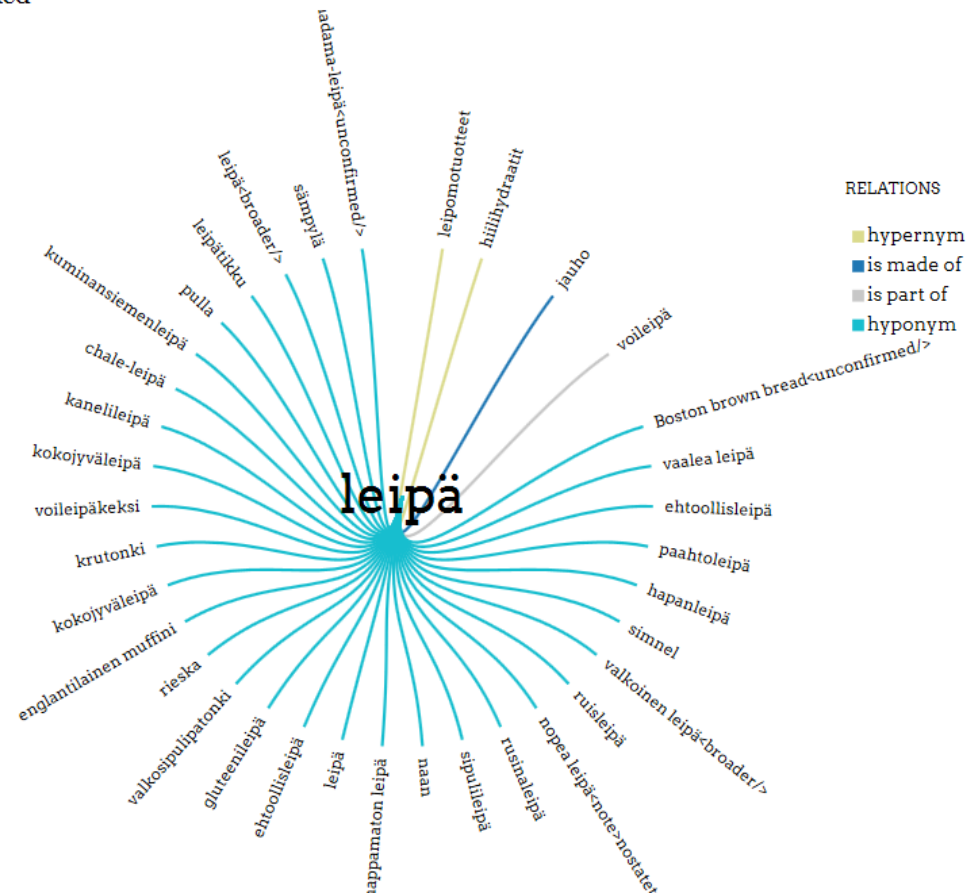
→ (Finnish) leipä : *food made from dough of flour or meal and usually raised with yeast or baking powder and then baked*

→ (Swedish) bröd :



**leipä**

food made from dough of flour or meal and usually raised with yeast or baking powder and then baked



## Example search in WordTies

Consider in more detail the sub-concepts of “leipä” (*bread*) and their definitions, such as the crisp bread “leipätikku” or the Indian bread “naan”

### Sub-concepts

*celebration of the Eucharist)*

- englantilainen muffini: round, raised muffin cooked on a griddle; usually split and toasted before being eaten
- gluteenileipä: bread made with gluten flour
- hapanleipä: made with a starter of a small amount of dough in which fermentation is active
- happamaton leipä and matsa: brittle flat bread eaten at Passover
- kanelileipä: bread flavored with cinnamon often containing raisins
- kokojyväleipä and tumma leipä: bread made with whole wheat flour
- kokojyväleipä: bread made with cracked wheat that has been ground fine
- krutonki, paahdettu leipäkuutio, and paistettu leipäkuutio: a small piece of toasted or fried bread; served in soup or salads
- kuminansiemenleipä: bread containing caraway seeds
- leipä and limppu: a shaped mass of baked bread that is usually sliced before eating
- leipä<broader/>: a rich currant cake or bun
- leipätikku: a crisp stick-shaped roll; often served with soup
- naan and naan-leipä: leavened bread baked in a clay oven in India; usually shaped like a teardrop
- nopea leipä<note>nostatettu soodalla</note>: breads made with a leavening agent that permits immediate baking



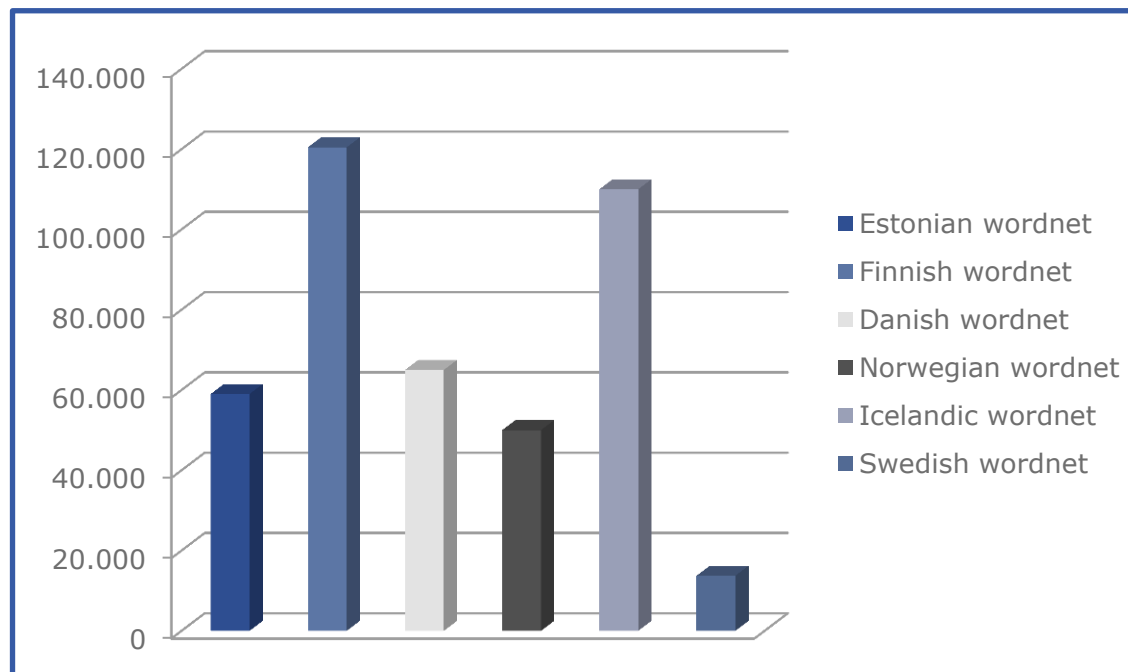
## Additional info about the current contents of WordTies



## Status of wordnets in the Nordic and Baltic countries

### Size:

- Estonian wordnet: 59,000 synsets
- Finnish wordnet: 120,449
- Danish wordnet: 66,308
- Swedish wordnet: 13,724
- Norwegian wordnets: 50,000
- Icelandic 'wordnet': 110,000



# Status of wordnets in the Nordic and Baltic countries

## Methodologies:

Expanded from PWN, monolingual extensions

- Estonian wordnet
- Finnish wordnet

Monolingually processed, linked to PWN Core

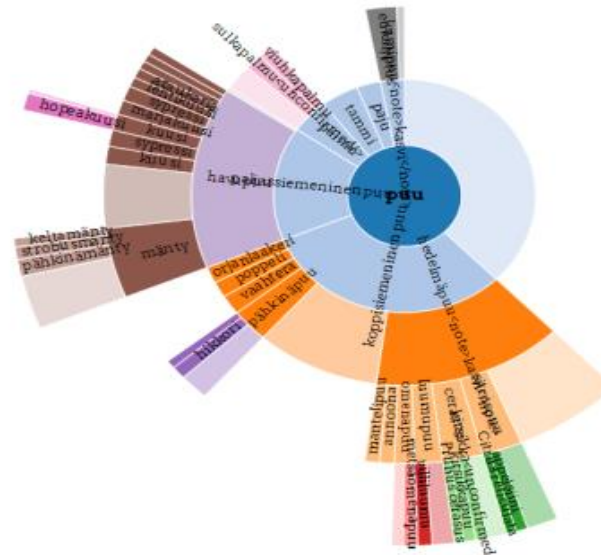
- Danish wordnet
- Swedish wordnet
- Icelandic wordnet

Expanded from the Danish wordnet, monolingual extensions

- Norwegian wordnets



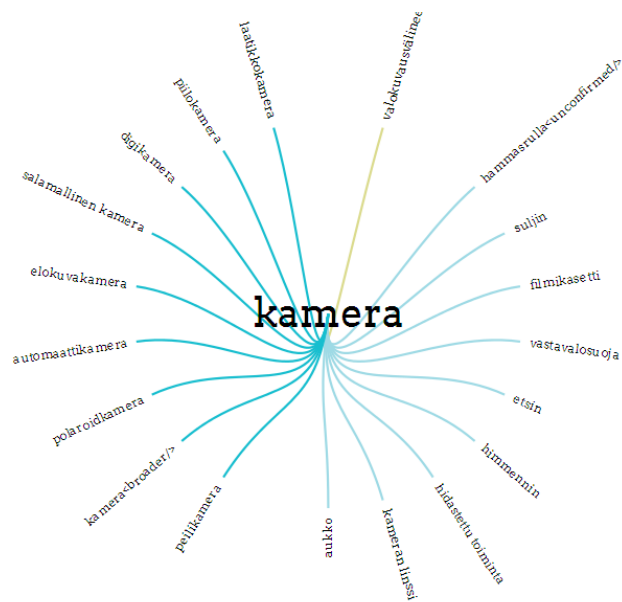
## Hyponym depth ('tree') in Danish and Finnish wordnets:





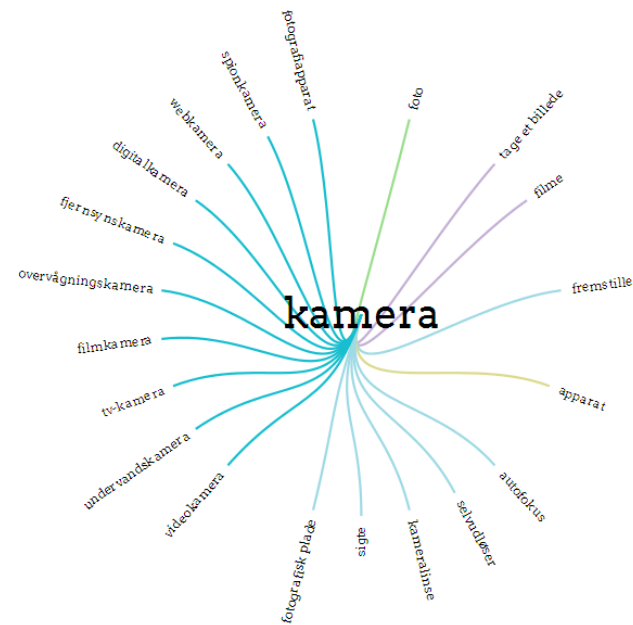
## Further comparison of selected wordnets

Relations per synset ('camera') in Danish and Finnish wordnets:



### RELATIONS

- hypernym
- has part
- hyponym



### RELATIONS

- domain
- used for
- made by
- hypernym
- has part
- hyponym



## Acknowledgements

The following wordnet developers participated in the task of aligning and evaluating the wordnets that are currently in WordTies: Lars Borin, Markus Forsberg, Neeme Kahusk, Krister Lindén, Jyrki Niemi, Niklas Nisbeth, Lars Nygaard, Heili Orav, Eirikur Rögnvaldsson, Mitchell Seaton, Kadri Vider, and Kaarlo Voionmaa.

