

# Workshop on the Collaborative Interlingual Index

- 09:00-10:00
  - Piek Vossen: Introducing the Global Wordnet Grid
  - John McCrae: ILI LMF and RDF
- 10:00-11:00
  - Hands-on: How to convert your wordnet to LMF and map to ILI
- 11:00-11:20 coffee break
- 11:20-13:20
  - Francis Bond: The Collaborative ILI
  - Christiane Fellbaum: How and when to add new concepts and how to define them

# The Global Wordnet Grid: a Collaborative Interlingual Index for interoperable wordnets

Piek Vossen, VU University Amsterdam, Netherlands

Francis Bond, Nanyang Technological University, Singapore

John McCrae, Insight Centre for Data Analytics at National  
University of Ireland, Galway

8th Global Wordnet Conference  
January, 27-30, 2016, Bucharest

# 25 years of wordnets

- Princeton WordNet since 1990 (1.5, 1.6, 2.0, 3.0, 3.1)
- EuroWordNet
- Multiwordnet, BalkaNet
- Indowordnet, Asian wordnet, African wordnet
- Open Multilingual Wordnet

# Status of the global net

- Built using different methods: merge, expand, manual or (semi-)automatic
- Different sets of relations were used
- Different interpretations of relations (with the same name)
- Different ways of defining synonyms (strict, loose)
- Different degrees of polysemy
- Differences in coverage

# Merging wordnets

middel

instrumentality

instrument

device

gereedschap, werktuig

instrument

boor

engine

apparaat, toestel

machine

machine

computer

motor

equipment

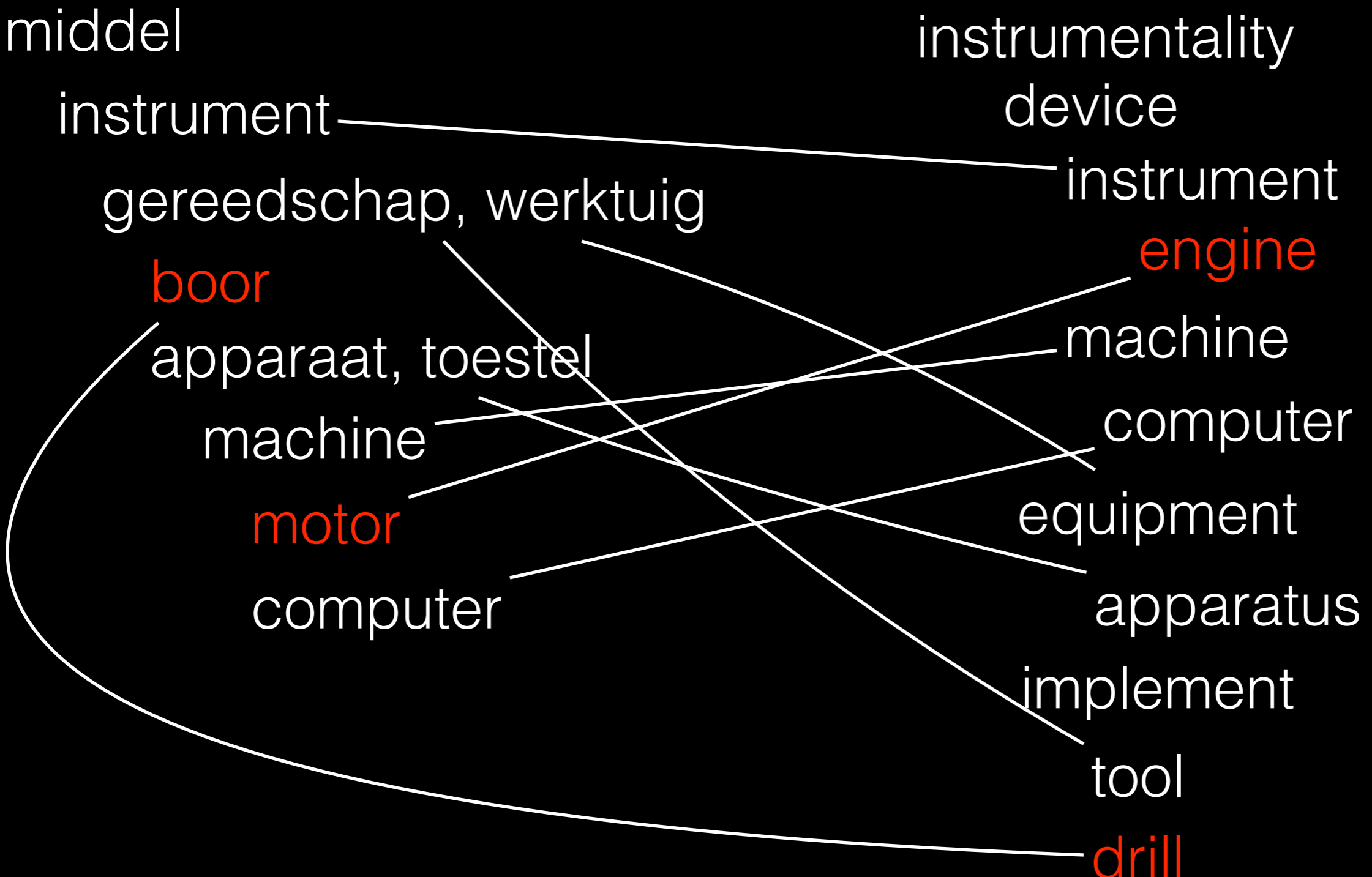
computer

apparatus

implement

tool

drill



# Status of the global net

- Linked to different versions of the Princeton WordNet as an Inter-Lingual-Index (ILI)
- Released in different formats
- Using different licenses
- Fixed Anglo-Saxon ILI: changes through Princeton
- No central hosting of the ILI and not available as LOD
- New version of PWN will mess up many wordnets and software systems that rely on it

# Open Multilingual Wordnet

- Open wordnets available for 33 languages, all linked to each other
- Automatically constructed data for 150 languages
- Postponed the question of how to link concepts across all languages by using PWN 3.0 as a de facto ILI, and dropping concepts that could not be linked

From a global net  
to  
the Global Wordnet Grid  
a platform for conceptual  
interoperability



# Global Wordnet Grid

- All open source wordnets linked to a single ILI
- Merge of concepts across all languages
- Princeton WordNet versions linked to ILI
- Adaptable by the open source wordnet community.
- Available as LOD and downloadable with open source license: CC-BY, CC-BY-SA.

# Motivation

- Achieving linguistic & conceptual interoperability:
- **Lexical** level:
  - Universals and idiosyncrasies in lexicalisation across languages
  - What is a word and a concept? What is relatedness & similarity?
- **Textual** level:
  - Interoperable understanding of **any** language by machines
  - E.g. NewsReader project: [www.newsreader-project.eu](http://www.newsreader-project.eu)

# NewsReader

## ILI-based event matching

nwr:ev21

```
a                                fn:Commerce_sell , ili:i32953 ,  
rdfs:label                       "verkopen" , "vendere" ;  
fn:Commerce_sell@Goods         dbp:Opel ;
```

GM said it was cancelling a deal to **sell** Opel to Magna...

GM ha comunicato l' annullamento di un accordo per **vendere** Opel a Magna...

zei GM dat het niet langer van plan was om Opel te **verkopen** aan Magna ...

# Should there be a limit?

- NO! we could even include adjective-noun or verb-object pairs or other phrases.
- What about productivity?
- Productivity and compositionality need to be observed.
- Cross-lingual lexicalisation determines value:
  - what is linked across languages works, what is not linked disappears eventually

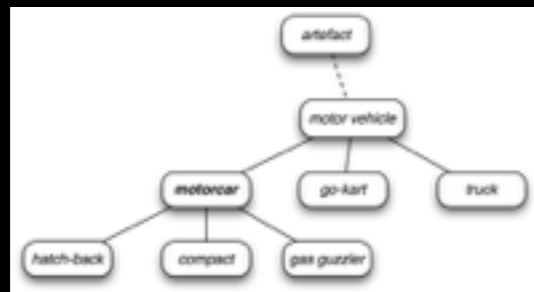
# What defines a concept?

- Keep it as simple as possible:
  - A unique URI (never to be deleted or modified)
  - Proper (!) English gloss is required
  - No part of speech
  - **LINKED** to a synset in at least one available open source wordnet through a owl:sameAs relation
  - **LINKED** to a hypernym synset in that wordnet or being a well-defined top : —> *no orphans!!*

# Protocol for changes

- Only providers of open source wordnets linked to the Grid can commit changes.
- How? By uploading an open source wordnet in the right format:
  - Commit new mapping file from a wordnet to existing ILI concepts
  - Commit proposals for new ILI concepts checked for eligibility
- Concepts can be voted for and commented on
- Concepts never disappear but can be ignored
- Concepts need to consistently defined in English using lengthy well-structured & explicit glosses
- Checks for duplicate glosses

# GWG ONION MODEL



**MODERATED**

VALIDATED

VOTED

Synsets  
not  
linked

Synsets  
not  
linked

ILI  
shared  
by all

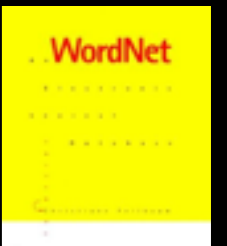
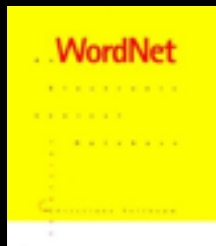
**ONTOLOGIZED**

ILI  
linked  
to 1

ILI  
shared  
by 2

ILI  
shared  
by 2

ILI  
linked  
to 1



# Infrastructure

- GWA website explains Grid, points to infrastructure and describes protocols and principles: <http://globalwordnet.org>
- Open Multilingual Wordnet repository: wordnets that feed the Grid, searchable, downloadable in standardised formats: <http://compling.hss.ntu.edu.sg/omw>
- Github project: <https://github.com/globalwordnet>
  - schemas, ILI and mappings to wordnets
- ILI SPARQL endpoint: <http://globalwordnet.org/sparql>



# Process

Wordnet builder

wn.vx.Imf

<http://compling.hss.ntu.edu.sg/omw>

ili-map-wnvx.ttl

OMW  
Search API  
Table download  
SPARQL endpoint

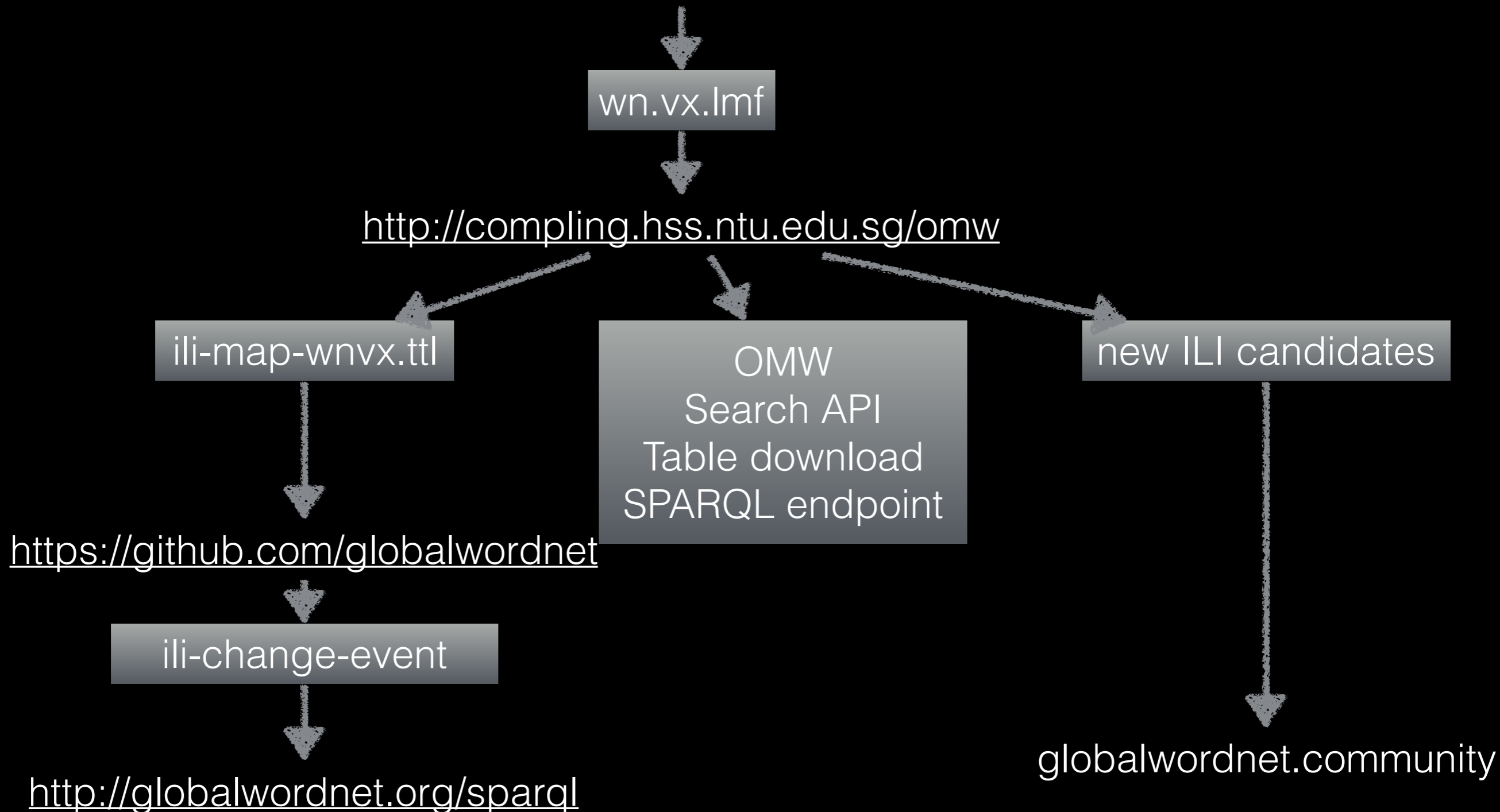
new ILI candidates

<https://github.com/globalwordnet>

ili-change-event

<http://globalwordnet.org/sparql>

globalwordnet.community



# ILI SPARQL DEMO

- <http://globalwordnet.org/ili/i78781>
- `select * where { <http://globalwordnet.org/ili/i78781> <http://www.w3.org/2002/07/owl#sameAs> ?m } LIMIT 100`
- `select * where { ?i <http://www.w3.org/2002/07/owl#sameAs> <http://wordnet-rdf.princeton.edu/wn30/14715356-n> } LIMIT 100`

In search of new ILI records  
in the Open Dutch Wordnet

# ODWN-1.0 (now 1.3)

<b>Open Dutch Wordnet 1.0</b>	<b>TOTAL</b>	<b>NOUNS</b>	<b>VERBS</b>
<b>Word forms</b>	57,602	50,255	7,347
<b>Lexical Units</b>	94,140	78,612	15,528
<b>Odwn Synsets with Dutch synonyms</b>	21,636	15,992	5,644
<b>Pwn Synsets with Dutch synonyms</b>	19,980	15,706	4,274
<b>Pwn Synsets without Dutch synonyms</b>	75,376	66,409	8,967
<b>Total</b>	116,992	98,107	18,885


1. Use more loose strategies for translations of lemmas:  
two sources were required in ODWN
2. Match hierarchical structures across wordnets, e.g.  
similarity measures across words and word senses
3. Match translations of glosses

# Matching strategies

- 21,636 ODWN synsets:
  - 7,489 match in another sense
  - 14,147 without match in any sense
- Google translate lemmas
- Google translate Dutch glosses

# Google translate lemmas

<b>LEMMAMATCHING</b>	<b>TOTAL</b>	<b>NOUNS</b>	<b>VERBS</b>
<b>MONOSEMOUS</b>	2,032	1,660	372
<b>DIRECT SYNONYM OR HYPERNYM</b>	2,455	1,939	516
<b>Leacock-Chodorow SIMILARITY</b>	335	334	1
<b>NOT ENTRY PWN</b>	5,067	4,236	831
<b>ENTRY BUT NO MATCH</b>	4,479	2,966	1,513
<b>TOTAL</b>	14,368	11,135	3,233



Try to match the translations of the glosses of co-hyponyms and co-hyponyms of hypernyms, e.g.

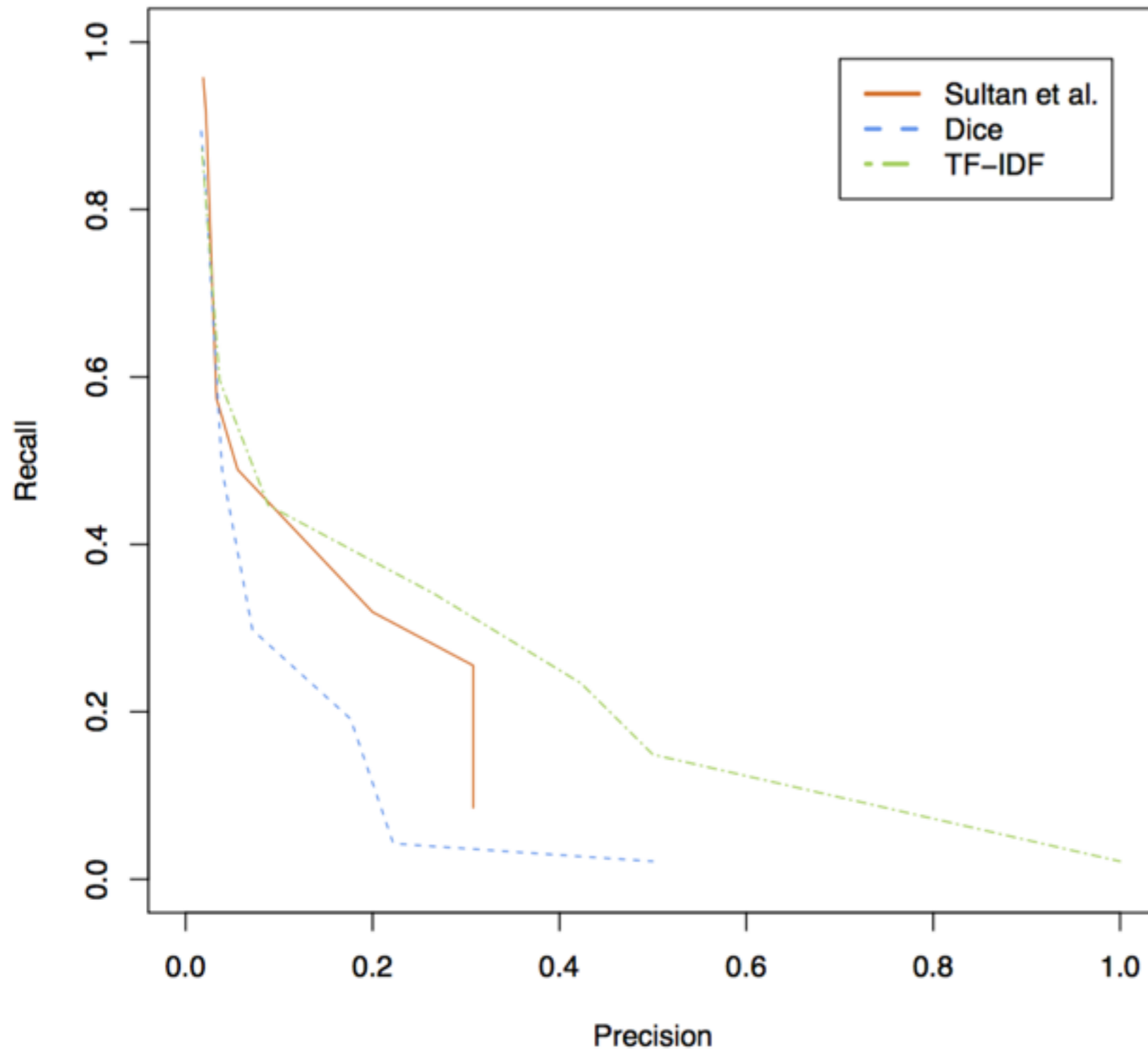
# Google translate glosses

bierbuik (beer belly):

gloss = “iemand met zo’n buik”

Google translate = “someone with such a belly”

<b>ODWN</b>	<b>DEFINITIONS</b>
<b>TOTAL</b>	<b>173,042</b>
<b>DUTCH</b>	<b>41,417</b>
<b>ENGLISH-PWN</b>	<b>95,190</b>
<b>ENGLISH-GoogleTrans.</b>	<b>36,435</b>



- Dice co-efficient
- Cosine of TF-IDF vectors of the glosses
- Alignment method by Sultan et al., 2014, best performance on the Semantic Textual Similarity Task at SemEval-2014

Figure 3: Precision-Recall curve of gloss matching between ODWN and PWN



# Main problem

- Quality and style of the glosses
  - afweersysteem (ODWN)
    - ‘afweer tegen ziektes’ (defense against diseases)
  - immune system (PWN)
    - a system (including the thymus and bone marrow and lymphoid tissues) that protects the body from foreign substances and pathogenic organisms by producing the immune response.

# Conclusion

- Most ODWN synsets without mapping can probably be mapped with sufficient effort
- Conservative to propose new concepts:
  - lemmas without Google translation
  - lemma translations that are not entries in PWN or with low WordNet similarity score
  - manually revised English glosses instead of translations
  - zero score for gloss similarity using different measures

# New ILI candidates

Dutch	English
perzikhuid	peach skin
kalfskotelet	veal chop
natuurramp	natural disaster
verwachtingspatroon	expectations
sluikreclame	product placement

Figure 2: ODWN entries not in PWN

## ZERO DICE SCORE

odwn-10-109994970-n [bierbuik][iemand met zo'n buik,someone with such a belly]  
odwn-10-109967773-v [instoppen][in bed leggen,lay in bed]  
odwn-10-109936203-n [kaalslag][het kappen van bomen,cutting down trees]  
odwn-10-109930274-n [anticlimax][tegenvaller bij hoge verwachtingen]  
odwn-10-109918239-n [vibrator][trillend toestel,vibrating device]  
odwn-10-109913187-n [zee-engte][vernaauwing in de zee door land]  
odwn-10-109890697-n [afweersysteem][afweer tegen ziektes,disease resistance]  
odwn-10-109871362-n [nieuwsvoorziening][verzorging v.h. nieuws]  
odwn-10-109865276-n [receptie][officiële ontvangst]  
odwn-10-109854072-v [bestellen][iemand laten komen,allow anyone]  
odwn-10-109837238-n [hamerstuk][vergaderpunt zonder discussie,meeting point without discussion]  
odwn-10-109834436-n [coupure][bedrag v.e. bankbiljet]  
odwn-10-109803922-v [ontlokken][bewering afdwingen,enforce allegation]  
odwn-10-109797015-v [afstrijken][strijkwerk afmaken,ironing finish]  
odwn-10-109796989-v [meedraaien][samen draaien,rotate together]  
odwn-10-109795703-n [imperialist][strever naar machtsuitbreiding,striver for power expansion]

## LOW DICE < 30

20 odwn-10-109974807-n [bootvluchteling][iemand per boot zijn land ontvlucht,someone boat fled country]  
20 eng-30-10017890-n [DP,displaced\_person,stateless\_person][a person forced to flee from home or country]  
29 odwn-10-109969533-n [schuimspaan][platte lepel met gaatjes,flat spoon with holes]  
5 eng-30-03623556-n [knife][gerei om mee te snijden,edge tool used as a cutting instrument; has a pointed blade with a sharp edge and a handle,parapherna  
19 eng-30-03361945-n [flatware][tableware that is relatively flat and fashioned as a single piece]  
17 eng-30-03025755-n [chopstick][one of a pair of slender sticks used as oriental tableware to eat food with]  
25 eng-30-04220945-n [silver\_plate][tableware that is plated with silver]  
25 eng-30-03445472-n [gold\_plate][tableware that is plated with gold]  
22 eng-30-02763714-n [awl][scherpe staaf om gaatjes te prikken,a pointed tool for marking surfaces or for punching small holes,sharp stick to poke holes]  
14 eng-30-03524574-n [hoe][landbouwwerktuig met gekromd blad om de grond los te hakken,bepaald tuingereedschap,a tool with a flat blade attached at right  
8 eng-30-03239726-n [drill][apparaat om gaten te maken,machinale boor,elektrische boormachine,a tool with a sharp point and cutting edges for making hole  
12 eng-30-03929202-n [pickaxe,pickax,pick][scherp en puntig werktuig,houweel met punt en beitel,a heavy iron tool with a wooden handle and a curved head t  
16 eng-30-04023249-n [puncher,punch][apparaat dat gaatjes maakt,apparaat dat kaarten ponst,a tool for making holes or indentations,device that makes holes  
6 eng-30-04051439-n [rammer][schroefpers,a tool for driving something with force,screw press]  
19 eng-30-02908217-n [brush][borstel met handvat,an implement that has hairs or bristles firmly set into a handle,brush with handle]  
5 eng-30-04050066-n [rake][tuingereedschap,a long-handled tool with a row of teeth at its head; used to move leaves or loosen soil,garden tools]  
18 eng-30-03403643-n [funnel][voorwerp met wijde hals en smalle tuit,a conically shaped utensil having a narrow tube at the small end; used to channel the  
16 eng-30-03454211-n [grappler,grappling\_iron,grapnel,grapple,grappling\_hook][a tool consisting of several hooks for grasping and holding; often thrown wi  
29 eng-30-03919430-n [muller,pestle,pounder][a heavy tool of stone or iron (usually with a flat base and a handle) that is used to grind and mix material

30 < = DICE < 70

50 odwn-10-109994127-n [flinter][zeer dun schijfje of stukje,very thin slice or shred]  
50 eng-30-09428741-n [section][a very thin slice (of tissue or mineral or other substance) for examination under a microscope]  
41 odwn-10-109993315-v [lachen][vrolijkheid uitdrukken,express mirth]  
35 eng-30-01067194-v [menace][express a threat either by an utterance or a gesture]  
41 eng-30-00929839-v [connote,imply][express or state indirectly]  
35 eng-30-01067380-v [beam][express with a beaming face or smile]  
37 eng-30-00032778-v [sneer][express through a scornful smile]  
41 eng-30-01067512-v [smile][express with a smile]  
41 eng-30-00928959-v [say][communicate or express nonverbally]  
41 odwn-10-109978071-n [varsity][roeiwedstrijd tussen studenten,rowing competition between students]  
41 eng-30-07468534-n [swimming\_meet,swim\_meet][a swimming competition between two or more teams]  
34 odwn-10-109975023-n [peillood][apparaat om diepte van water te meten,device to measure depth of water]  
34 eng-30-04260192-n [sonic\_depth\_finder,fathometer][depth finder for determining depth of water or a submerged object by means of sound]  
37 odwn-10-109960175-n [breedtecirkel][denkbeeldige cirkel over de aarde,imaginary circle the earth]  
37 eng-30-13873849-n [circlet][a small circle]  
32 eng-30-05102889-n [osculating\_circle,circle\_of\_curvature][the circle that touches a curve (on the concave side) and whose radius is the same as that of the curve]  
37 odwn-10-109950849-v [stationeren][een standplaats geven,give a stand]  
37 eng-30-01517515-v [upend][set, turn, or stand on end]  
37 eng-30-01515196-v [launch,set\_in\_motion][get going; give impetus to]  
33 odwn-10-109949756-v [meerijden][met anderen rijden,others ride]  
33 eng-30-02376958-v [interact][act together or towards others or with others]  
36 odwn-10-109945969-n [matje][lang nekhaar bij een verder kort kapsel,long neck hair with a further short haircut]  
36 eng-30-05259109-n [bob][a hair style for women and children; a short haircut all around]

DICE  $\geq$  70

83 odwn-10-109697118-n [brandschade][schade door brand,damage by fire]  
83 eng-30-00403911-n [burn][damage inflicted by fire]  
75 odwn-10-109616370-v [opbieden][hoger bieden,bid higher]  
75 eng-30-02299801-v [outbid][bid higher than others]  
71 odwn-10-109571923-n [wikkel][papiertje waarin iets verpakt wordt,paper in which something is wrapped]  
71 eng-30-04605726-n [wrap,wrapping,wrapper][the covering (usually paper or cellophane) in which something is wrapped]  
83 odwn-10-109187549-n [schuimplastic][bepaalde kunststof]  
83 eng-30-03979377-n [polyester][bepaalde harde kunststof,any of a large class of synthetic fabrics,certain hard plastics]  
73 odwn-10-109076330-n [hit][klein soort paard,kind of small horse]  
73 eng-30-02382437-n [pony][klein soort paard,any of various breeds of small gentle horses usually less than five feet high]  
83 odwn-10-108712975-n [rammelaar][mannetje van het konijn,male rabbit]  
83 eng-30-02324514-n [lapin][castrated male rabbit]  
83 odwn-10-108646433-n [rotswand][steile zijde v.e. rots]  
83 eng-30-09246464-n [cliff,drop,drop-off][steile rots,a steep high face of rock,cliff]  
76 odwn-10-108287212-v [uitknippen][vormen met een schaar,shapes with scissors]  
76 eng-30-01248950-v [scissor][met een schaar,cut with or as if with scissors,with scissors]  
72 odwn-10-107859308-n [Maagd][één v.d. twaalf sterrenbeelden,one of the Twelve Constellations]  
72 eng-30-09232687-n [cancer][één v.d. twaalf sterrenbeelden,a small zodiacal constellation in the northern hemisphere; bet  
77 odwn-10-107748764-v [omklappen][van stand veranderen,of state change]  
77 eng-30-01831531-v [move][van stand veranderen,move so as to change position, perform a nontranslational motion,of state  
75 odwn-10-107646993-v [afgieten][kookwater weggieten,cooking water pour out]  
75 eng-30-02070296-v [pour\_out,decant,pour][pour out]  
75 eng-30-02069788-v [pour\_out,effuse][pour out]