Transmute WordNet into Lexical system

One Lexicon, Two Structures: So What Gives?
Lexical System structure

Nodes
- Lexemes
- Idioms
- Linguistic cliches

Arcs
- Lexical functions links
- Phraseology links
- Copolysemy links

Lexicographic article associated with each lexical unit: GC, DF, GP, LF, EX, PH

sollinge@atilf.fr
One Lexicon, Two Structures: So What Gives? 29 January 2014
Lexical System structure

Nodes
- Lexemes
- Idioms
- Linguistic cliches

Arcs
- Lexical functions links
- Phraseology links
- Copolysemy links

Lexicographic article associated with each lexical unit:
- GC, DF, GP, LF, EX, PH

Confidence index
Lexical system by example: fr-LN

**Nodes**
- 21,507 nodes
  - 55% N
  - 21% V
  - 16% Adj
  - 0.5% Adv
  - 16% single nodes

**Arcs**
- 39,777 arcs
  - 88% LF links
    - 1/5 syntagmatic
    - 4/5 paradigmatic
  - 10% phraseology links
  - 2% copolysemy links
- 39 loops
- 515 multiples
- 18,530 mutuals
Lexical system by example: fr-LN

Nodes
- 21,507 nodes
  - 55% N
  - 21% V
  - 16% Adj
  - 0.5% Adv
- 16% single nodes

Arcs
- 39,777 arcs
  - 88% LF links
    - 1/5 syntagmatic
    - 4/5 paradigmatic
  - 10% phraseology links
  - 2% copolysemy links
- 39 loops
- 515 multiples
- 18,530 mutuals

Workforce
- 10 full-time lexicographers since June 2011
Evolution of fr-LN

Tends to become Hierarchical Small World Networks

sollinge@atilf.fr

One Lexicon, Two Structures: So What Gives?

29 January 2014
Evolution of fr-LN

Tends to become Hierarchical Small World Network
Why "transmute" WordNet into LN system?

Start work on the English Lexicon

Explore structural behavior of LN
How to ”transmute” WordNet 3.0 into en-LN ?

Nomenclature

One Lexicon, Two Structures: So What Gives?
How to ”transmute” WordNet 3.0 into en-LN ?

Nomenclature

wn_s.pl
wn_g.pl
wn_sk.pl
wn_syntax.pl
How to "transmute" WordNet 3.0 into en-LN?

Nomenclature

156,584 en-LN vocables

| idVoc | form | subscript |
|-------|------|-----------|

subscript : null, N, V, Adj, Adv

wn_s.pl
wn_g.pl
wn_sk.pl
wn_syntax.pl
How to "transmute" WordNet 3.0 into en-LN?

Nomenclature

156,584 en-LN vocables

| idVoc | form | subscript |
|-------|------|-----------|
|       |      | null,N,V,Adj,Adv |

206,976 en-LN senses

| idSense | idVoc | sense number | gloss |
|---------|------|--------------|-------|
How to "transmute" WordNet into en-LN?

Relations

wn_s.pl
wn_hyp.pl
wn_ins.pl
wn_sim.pl
wn_mm.pl
wn_ms.pl
wn_mp.pl
wn_der.pl
wn_at.pl
wn_per.pl
wn_cs.pl
wn_ant.pl
wn_ppl.pl
How to "transmute" WordNet into en-LN?

Relations

wn_s.pl
wn_hyp.pl
wn_ins.pl
wn_sim.pl
wn_mm.pl
wn_ms.pl
wn_mp.pl
wn_der.pl
wn_at.pl
wn_per.pl
wn_cs.pl
wn_ant.pl
wn_ppl.pl

946,208 en-LN LF links

| idLF | idSenseSource | idSenseTarget |
|------|---------------|---------------|

11 pre-existent LF
1 new LF: Unspecified derivative
How to ”transmute” WordNet into en-LN ?

Lexicographic article

wn_s.pl
wn_fr.pl
How to "transmute" WordNet into en-LN?

Lexicographic article

en-LN senses GC

| idSense | idGC | form? |
|---------|------|-------|

wn_s.pl
wn_fr.pl
How to ”transmute” WordNet into en-LN ?

Lexicographic article

en-LN senses GC

wn_s.pl
wn_fr.pl

en-LN senses GP comments

idSense | idGC | form?

idSense | subcategorization frames
Blue-print of an English lexical network

- Adjust actual links
- Increase connectivity
  - LF
  - Co-polysemy
  - Phraselogy
- Complete lexicographic descriptions

Tools

En-LN is compatible with our lexicographic editor
Thanks for your attention
Pedigree of th fr-/en-LNs (global)

|               | fr-LN   | en-LN   |
|---------------|---------|---------|
| n             | 21,507  | 206,976 |
| m             | 39,777  | 946,208 |
| $<k>$         | 3.6406  | 5.9029  |
| Directed      | true    | true    |
| Mutuals       | 18,530  | 942,795 |
| Loops         | 39      | 1       |
| Single        | 3,404   | 19,756  |
| Multiples     | 515     | 124     |
| ncc           | 14,013  | 34,342  |
| C             | 0.1270  | 0.1031  |
| Out degree distribution |     |         |
| a             | -2.0454 | -1.8479 |
| $r^2$         | 0.9541  | 0.8453  |

C Random

|               | C Random |
|---------------|----------|
| fr-LN         | 0.00017  |
| en-LN         | 0.00004  |
# Pedigree of th fr-/en-LNs (LCC)

|       | fr-LN  | en-LN  |
|-------|--------|--------|
| n_lcc | 2,741  | 144,294|
| m_lcc | 9,979  | 851,748|
| C_lcc | 0.3225 | 0.0980 |
| L_lcc | 12.0383| 10.1479|

|       | logn/ log logn |
|-------|----------------|
| fr-LN | 6.4105         |
| en-LN | 7.240          |
Wordnet Synset

s(100478262,1,'soccer',n,1,0).
s(100478262,2,'association football',n,1,0).

en-LN LF

\[ \text{Syn}(\text{SOCCER}) = \text{ASSOCIATION FOOTBALL} \]
\[ \text{Syn}(\text{ASSOCIATION FOOTBALL}) = \text{SOCCER} \]
Cf

Wordnet Synset

\[\text{sim}(301092142,301092572).\]
\[\text{s}(301092142,1,\text{‘malfunctioning’,a,1,0}).\]
\[\text{s}(301092142,2,\text{‘nonfunctional’,a,2,0}).\]
\[\text{s}(301092572,1,\text{‘bad’,s,14,0}).\]

en-LN LF

\[
\begin{align*}
\text{Cf}(\text{MALFUNCTIONING}) &= \text{BAD}_{\text{Adj}} 14 \\
\text{Cf}(\text{NONFUNCTIONAL}) &= \text{BAD}_{\text{Adj}} 14 \\
\text{Cf}(\text{BAD}_{\text{Adj}} 14) &= \text{MALFUNCTIONING, NONFUNCTIONAL}
\end{align*}
\]
Hypo - Gener

Wordnet Synset

hyp(110129825,110787470).

s(110129825,1,'girl',n,1,80).

s(110787470,1,'woman',n,1,143).

s(110787470,2,'adult female',n,1,0).

s(110787470) more than 15 times in wn_hyp.pl

en-LN LF

**Gener**(GIRL 1) = WOMAN 1, ADULT FEMALE

**Hypo**(WOMAN 1) = GIRL 1

**Hypo**(ADULT FEMALE) = GIRL 1
Hypo - Gener

Wordnet Synset

ins(109012735,108524735).
s(109012735,1,'Tartu',n,1,0).
s(108524735,1,'city',n,1,103).
s(108524735,2,'metropolis',n,1,7).
s(108524735,3,'urban center',n,1,2).

en-LN LF

Gener(Tartu) = city 1,metropolis 1,urban center
Hypo(city 1) = Tartu
Hypo(metropolis 1) = Tartu
Hypo(urban center) = Tartu
Sing (team\textsubscript{N} 1) = stringer 1
Sing (squad 2) = stringer 1
Mult (stringer 1) = team\textsubscript{N} 1, squad 2
Wordnet Synset

ms(101896844,103266749).

s(101896844,1,'eiderdown',n,2,0).
s(103266749,1,'eiderdown',n,1,0).
s(103266749,2,'duvet',n,1,0).
s(103266749,3,'continental quilt',n,1,0).

en-LN LF

Mero(eiderdown 1) = eiderdown 2
Mero(duvet) = eiderdown 2
Mero(continental quilt) = eiderdown 2
Wordnet Synset

mp(101896844,101853195).
s(101896844,1,'eiderdown',n,2,0).
s(101853195,1,'eider',n,1,0).
s(101853195,2,'eider duck',n,1,0).

en-LN LF

Mero(EIDER) = EIDERDOWN 2
Mero(EIDER DUCK) = EIDERDOWN 2
Holo(EIDERDOWN 2) = EIDER,EIDER DUCK
Unspecified derivative

Wordnet Synset
at(302410393,105103072).
s(302410393,1,’thick’,a,1,25).
s(105103072,1,’thickness’,n,1,4).

en-LN LF
Unspecified derivative(THICK\textsubscript{A}dj) = THICKNESS 1
Unspecified derivative(THICKNESS 1) = THICK\textsubscript{A}dj

Also
+ derivational morphology pointers
+ pertainym pointers
Wordnet Synset

\[ \text{at}(202380571,202379753). \]
\[ s(202380571,1,'pension off',v,1,0). \]
\[ s(202379753,1,'retire',v,1,18). \]

en-LN LF

\[ \text{Caus}(\text{RETIRE 1}) = //\text{PENSION OFF 1} \]
Wordnet Synset

ant(100080743,1,100080968,1).
ant(100080968,1,100080743,1).

s(100080743,1,'call option',n,2,0).
s(100080968,1,'put option',n,2,0).
s(100080968,2,'put',n,1,0).

en-LN LF

\[ \text{Anti} \cap (\text{CALL OPTION 2}) = \text{PUT OPTION 2} \]
\[ \text{Anti} \cap (\text{PUT OPTION 2}) = \text{CALL OPTION 2} \]
Wordnet Synset

ppl(303152480,1,200991683,1).

s(303152480,1,'posted',a,1,0).
s(200991683,1,'post',v,2,3).

en-LN LF

\( A_2(\text{POST}^\nu 2) = \text{POSTED} \)
Duplicates and Co

- 5,580 word senses
- 10 derivational morphology pointers
- 3,222 pertainym pointers
- 710 antonymous pairs

- s(301380267,1,'aerial',s).
- **wrongly coded**
- no corresponding sense in our database
- 2 derivational morphology pointers lost
Troponymy between talk and whisper

en-LN LF

\textbf{Gener}(\textsc{whisper}) = \textsc{talk}_V 2, \ldots \\
\textbf{Hypo}(\textsc{talk}_V 2) = \textsc{whisper} \\
\ldots

corrected en-LN LF

\textbf{Syn}_\subseteq(\textsc{whisper}) = \textsc{talk}_V 2, \ldots \\
\textbf{Syn}_\supset(\textsc{talk}_V 2) = \textsc{whisper} \\
\textbf{AntiMagn}(\textsc{talk}_V 2) = \textsc{whisper}