



Semantic Mirrors

*Thesaurus
derivation from
parallel corpora*

Helge Dyvik
University of Bergen

Picasso

Overview

1. *Meaning, semantic representations and translation*
2. *Sense individuation based on a parallel corpus*
3. *Deriving semantic wordnets from a parallel corpus*
4. *A toy semantic field*
5. *Evaluating the results; future perspectives*

Princeton WordNet online:

Results for "Synonyms, ordered by estimated frequency" search of noun "society"

4 senses of *society*

Sense 1

society -- (an extended social group having a distinctive cultural and economic organization)

=> social group -- (people sharing some social relation)

Sense 2

club, society, guild, gild, lodge, order -- (a formal association of people with similar interests; "he joined a golf club"; "they formed a small lunch society"; "men from the fraternal order will staff the soup kitchen today")

=> association -- (a formal organization of people or groups of people; "he joined the Modern Language Association")

Sense 3

company, companionship, fellowship, society -- (the state of being with someone; "he missed their company"; "he enjoyed the society of his friends")

=> friendship, friendly relationship -- (the state of being friends)

Sense 4

society, high society, beau monde, smart set, bon ton -- (the fashionable elite)

=> elite -- (a group or class of persons enjoying superior intellectual or social or economic status)

Princeton WordNet online:

Results for "Hypernyms (this is a kind of...)" search of noun "society"

4 senses of *society*

Sense 1

society -- (an extended social group having a distinctive cultural and economic organization)

=> social group -- (people sharing some social relation)

=> group, grouping -- (any number of entities (members) considered as a unit)

Sense 2

club, society, guild, gild, lodge, order -- (a formal association of people with similar interests; "he joined a golf club"; "they formed a small lunch society"; "men from the fraternal order will staff the soup kitchen today")

=> association -- (a formal organization of people or groups of people; "he joined the Modern Language Association")

=> organization, organisation -- (a group of people who work together)

=> social group -- (people sharing some social relation)

=> group, grouping -- (any number of entities (members) considered as a unit)

Sense 3

company, companionship, fellowship, society -- (the state of being with someone; "he missed their company"; "he enjoyed the society of his friends")

=> friendship, friendly relationship -- (the state of being friends)

=> relationship -- (a state involving mutual dealings between people or parties or countries)

=> state -- (the way something is with respect to its main attributes; "the current state of knowledge"; "his state of health"; "in a weak financial state")

Princeton WordNet online:

Sense 4

society, high society, beau monde, smart set, bon ton -- (the fashionable elite)

=> elite -- (a group or class of persons enjoying superior intellectual or social or economic status)

=> upper class, upper crust -- (the class occupying the highest position in the social hierarchy)

=> class, social class, socio-economic class -- (people having the same social or economic status; "the working class"; "an emerging professional class")

=> people -- ((plural) any group of human beings (men or women or children) collectively; "old people"; "there were at least 200 people in the audience")

=> group, grouping -- (any number of entities (members) considered as a unit)

Semantic field

A “meaning continuum” which the words in a language carve up in their own particular way.

(Examples: colour terms, emotional terms)

Concepts in a semantic field

animal

pet

feline

canine

cat

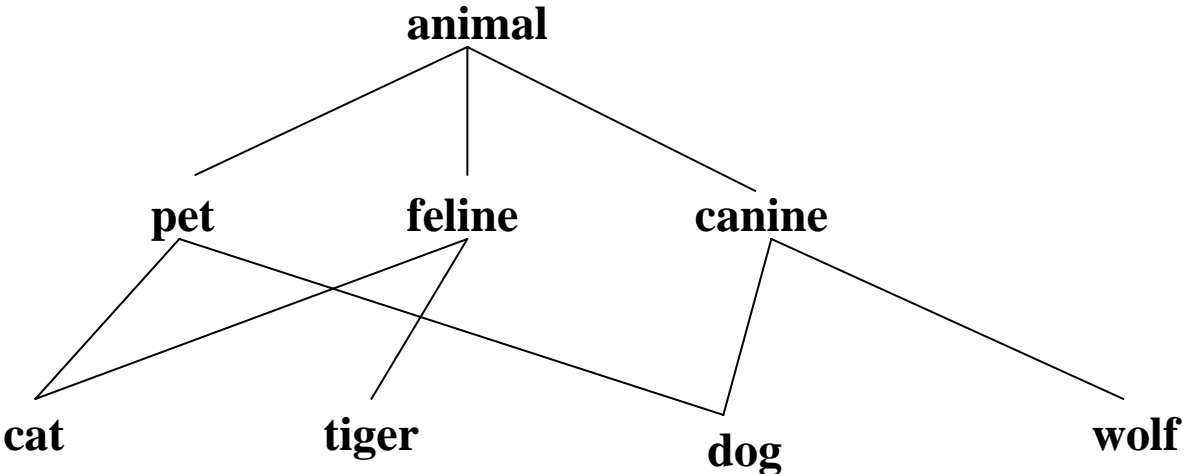
tiger

dog

wolf

Concepts in a semantic field

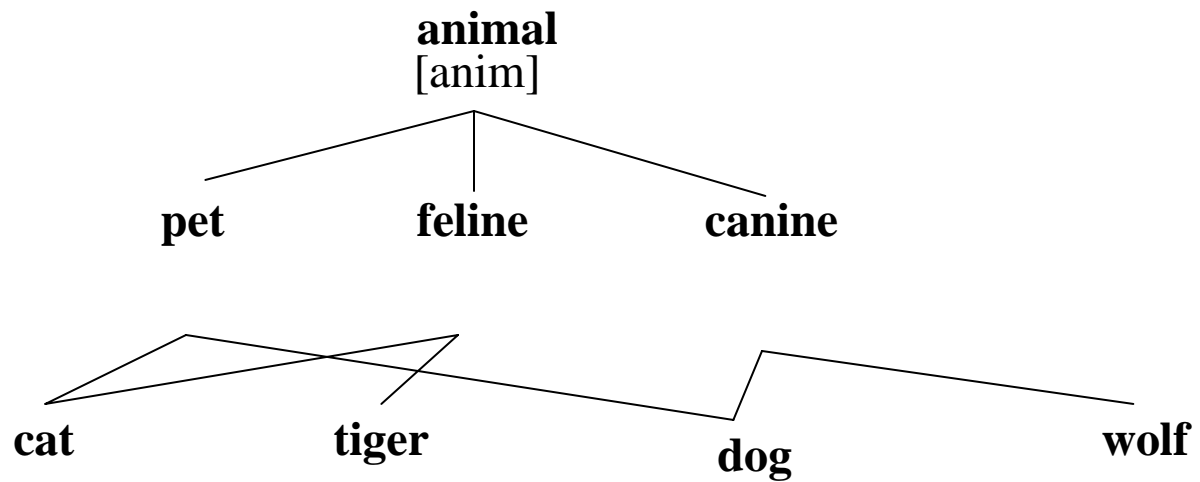
Lattice structure



Concepts in a semantic field

Lattice structure

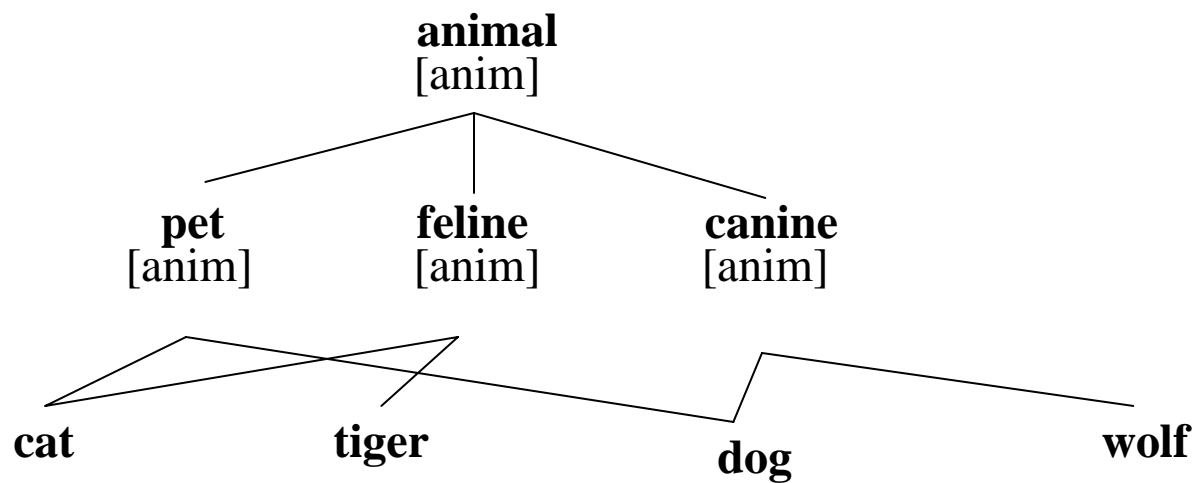
Feature analysis



Concepts in a semantic field

Lattice structure

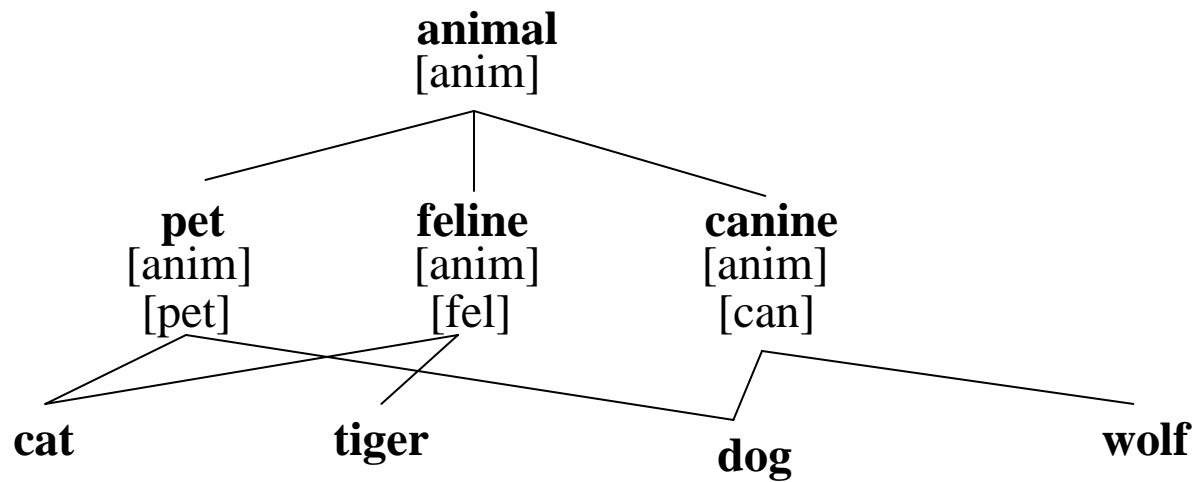
Feature analysis



Concepts in a semantic field

Lattice structure

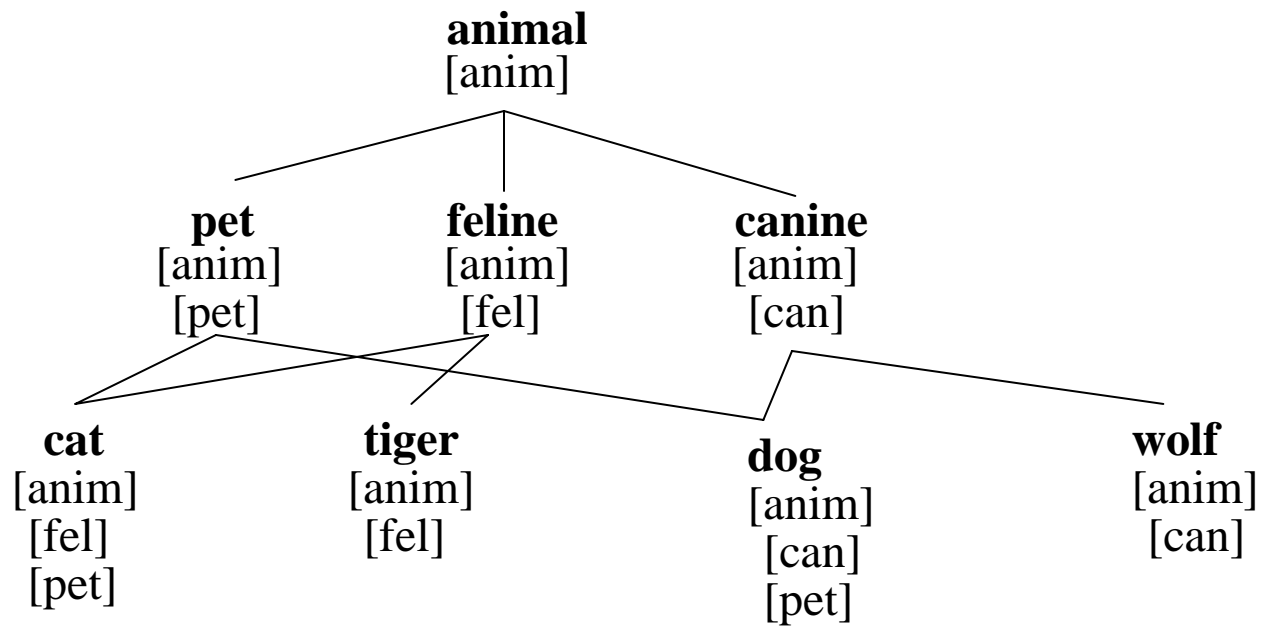
Feature analysis



Concepts in a semantic field

Lattice structure

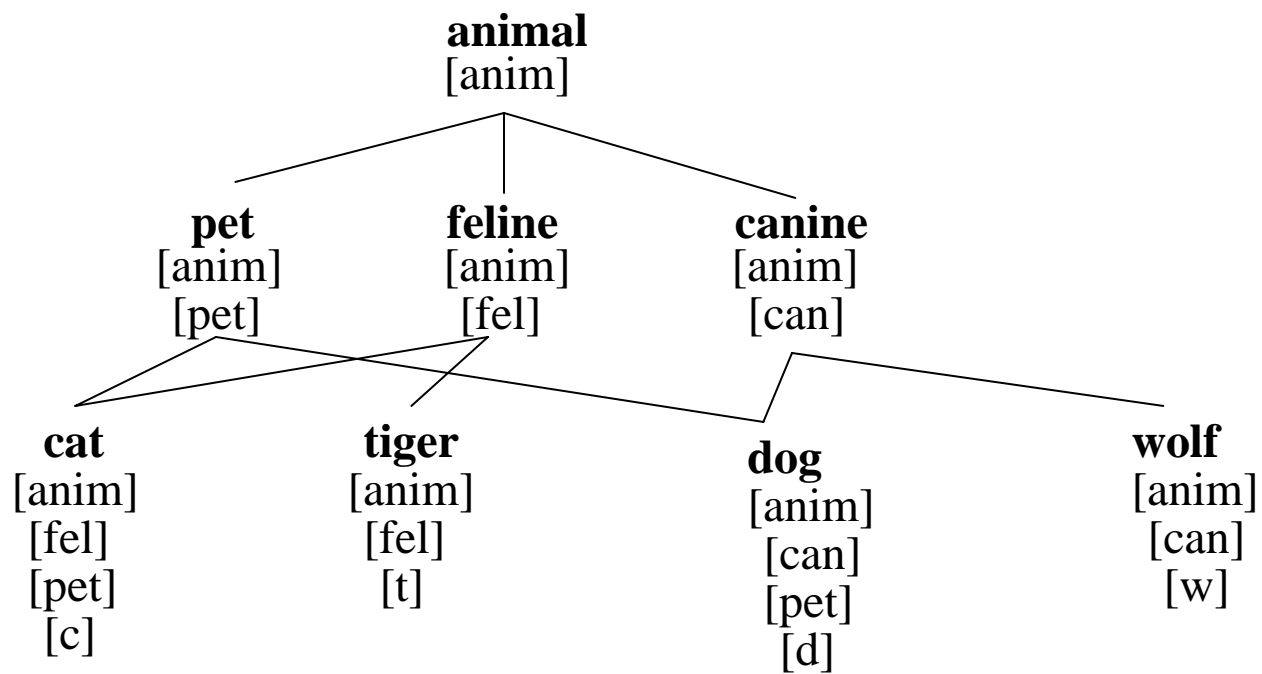
Feature analysis



Concepts in a semantic field

Lattice structure

Feature analysis



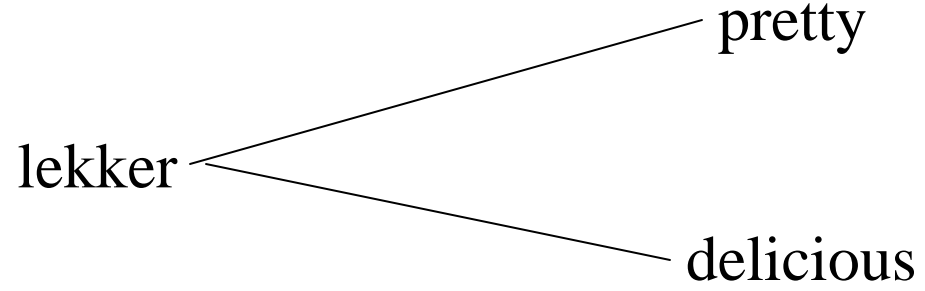
Norw

Eng

lekker

pretty

delicious



Norw

Eng

lekker

pretty

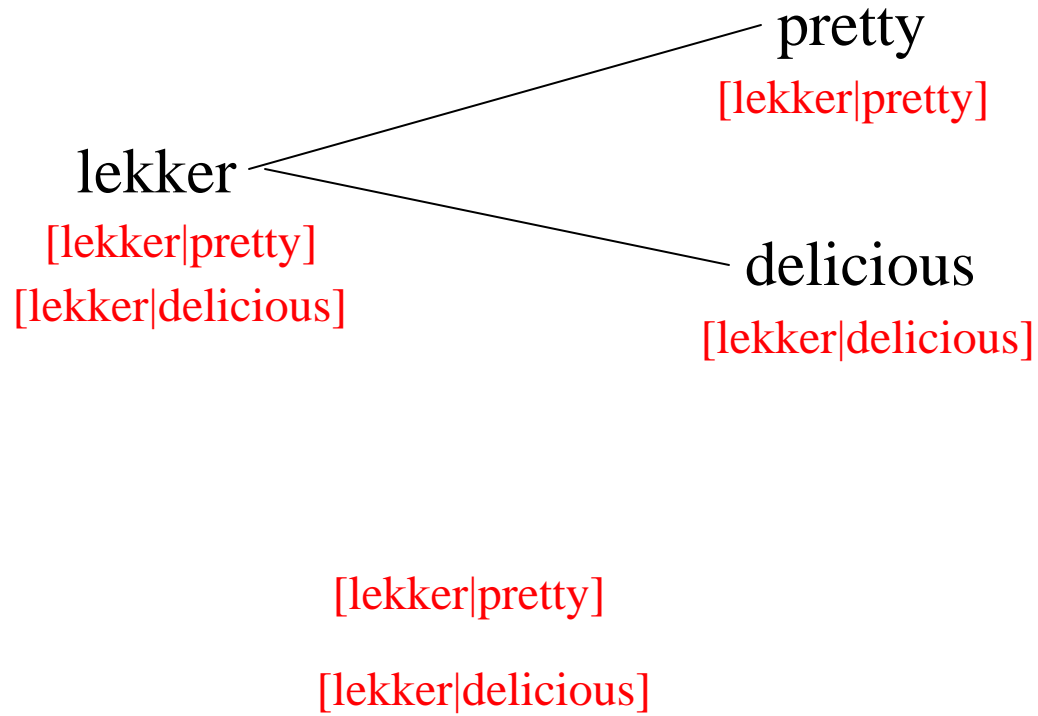
delicious

[lekker|pretty]

[lekker|delicious]

Norw

Eng

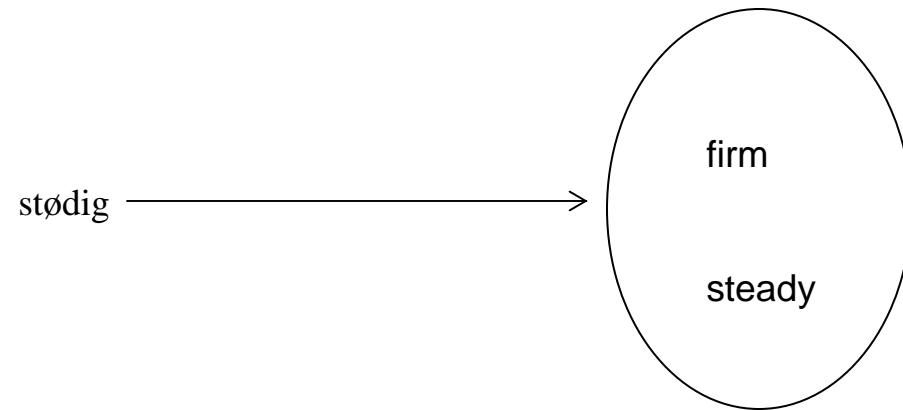


Literal or non-literal translation?

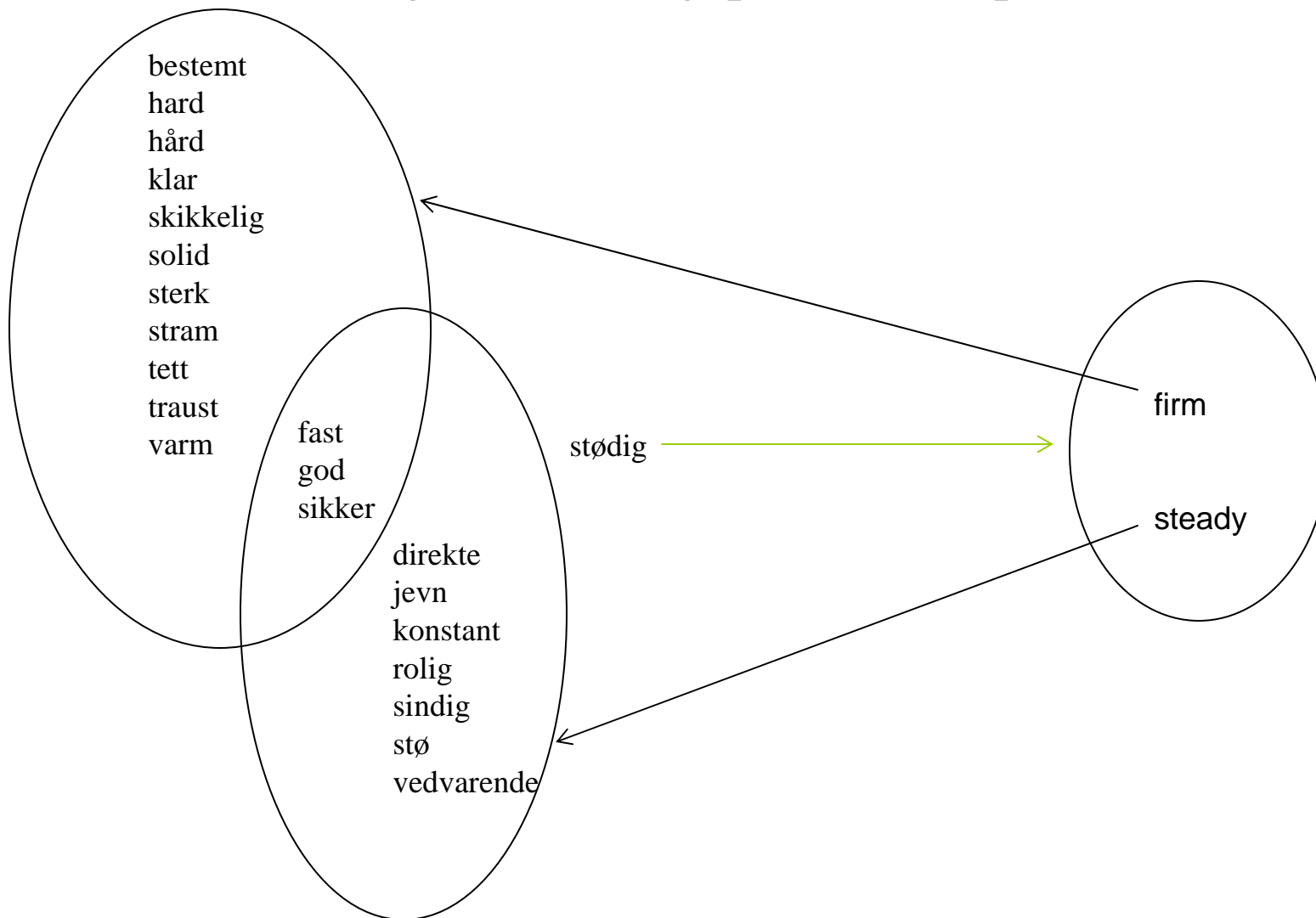
(from the EEA-agreement / EØS-avtalen)

carriers ————— utøvere av transportvirksomhet
(performers of transport activities)

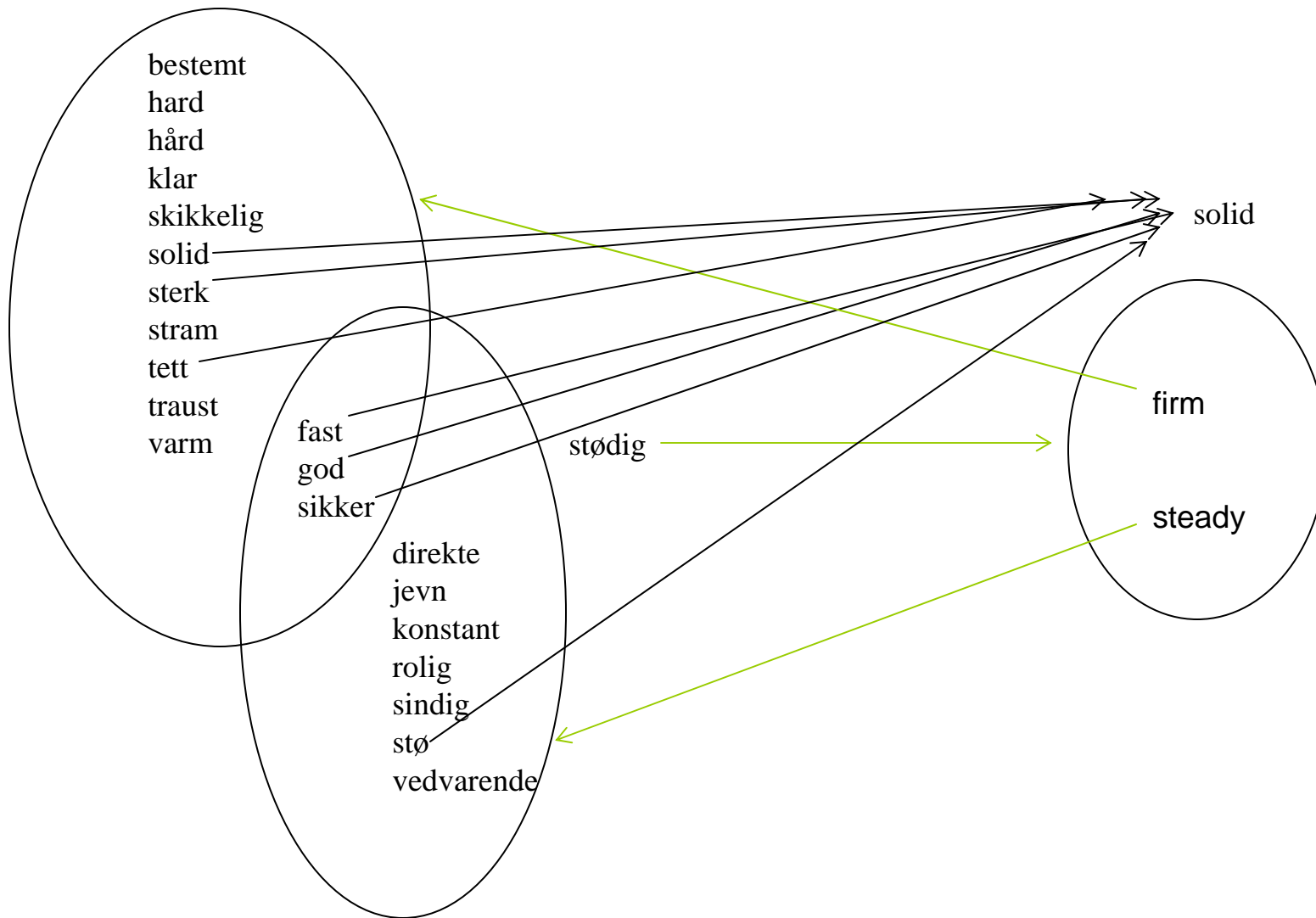
Filling accidental gaps in the corpus:



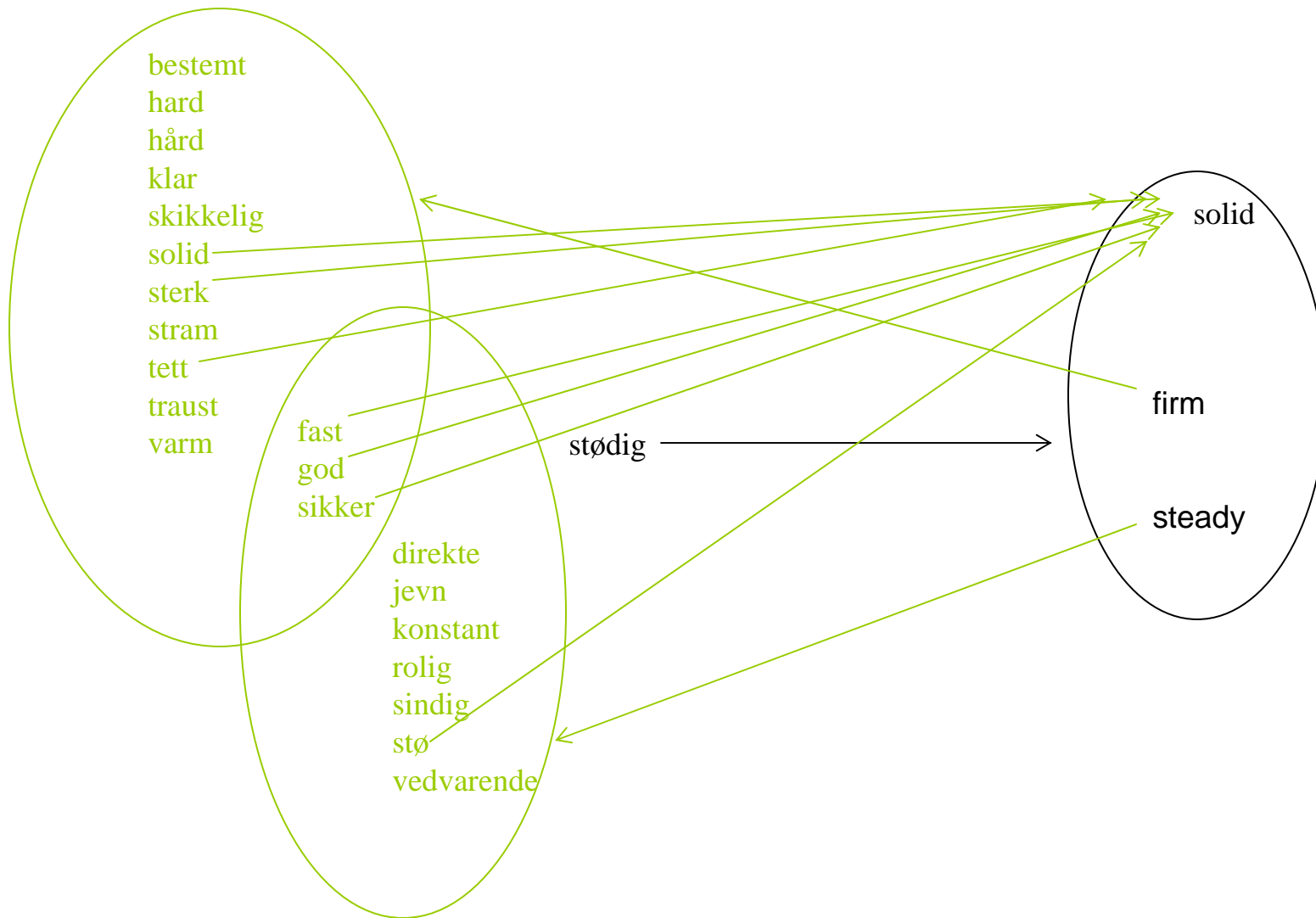
Filling accidental gaps in the corpus:



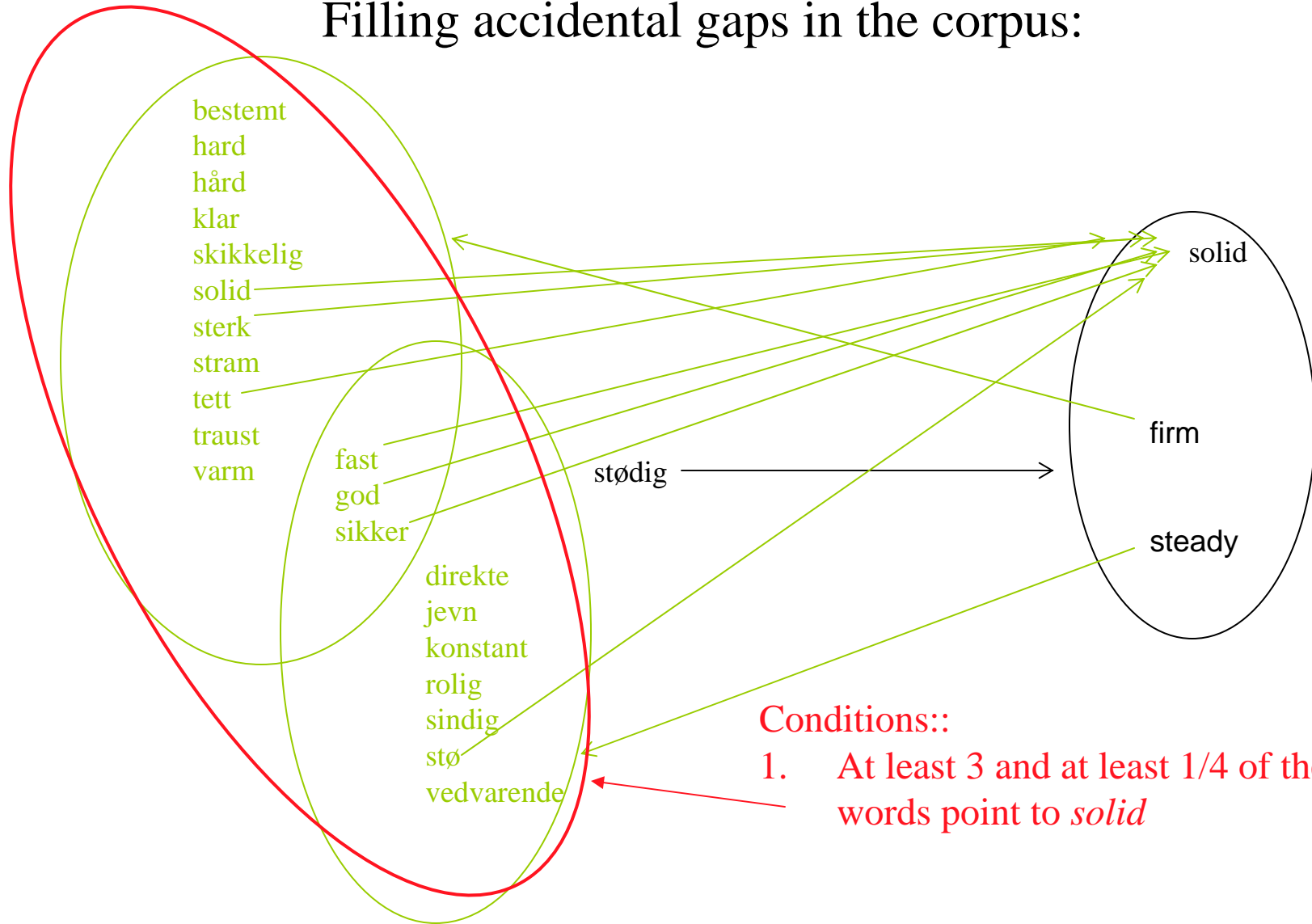
Filling accidental gaps in the corpus:



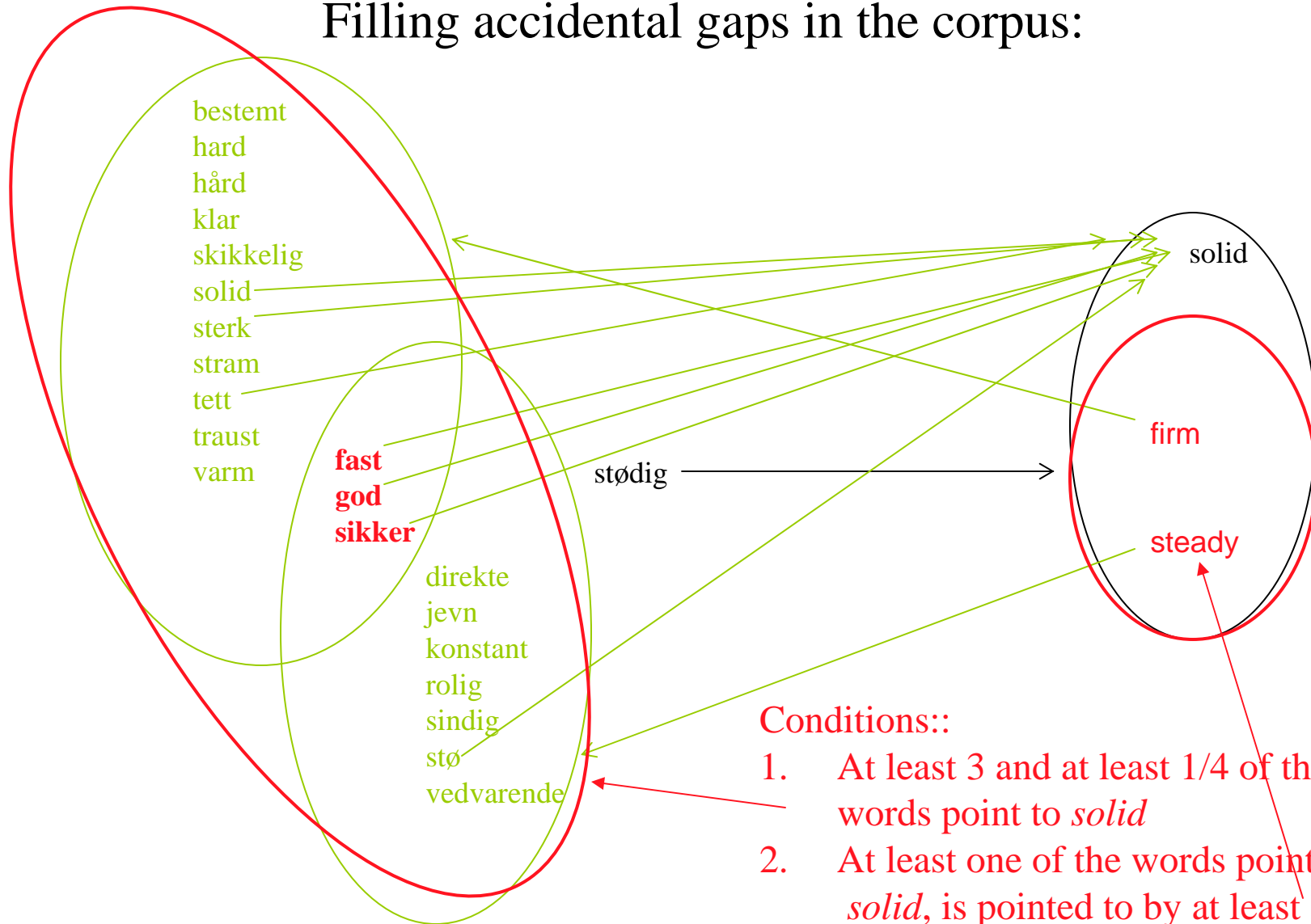
Filling accidental gaps in the corpus:



Filling accidental gaps in the corpus:



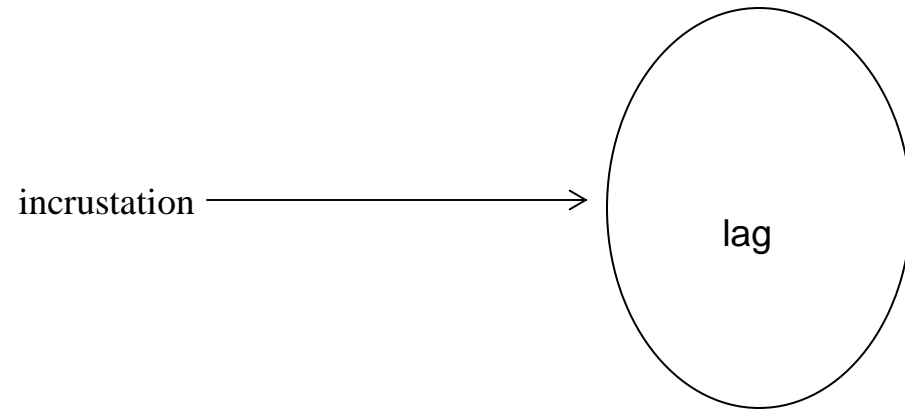
Filling accidental gaps in the corpus:



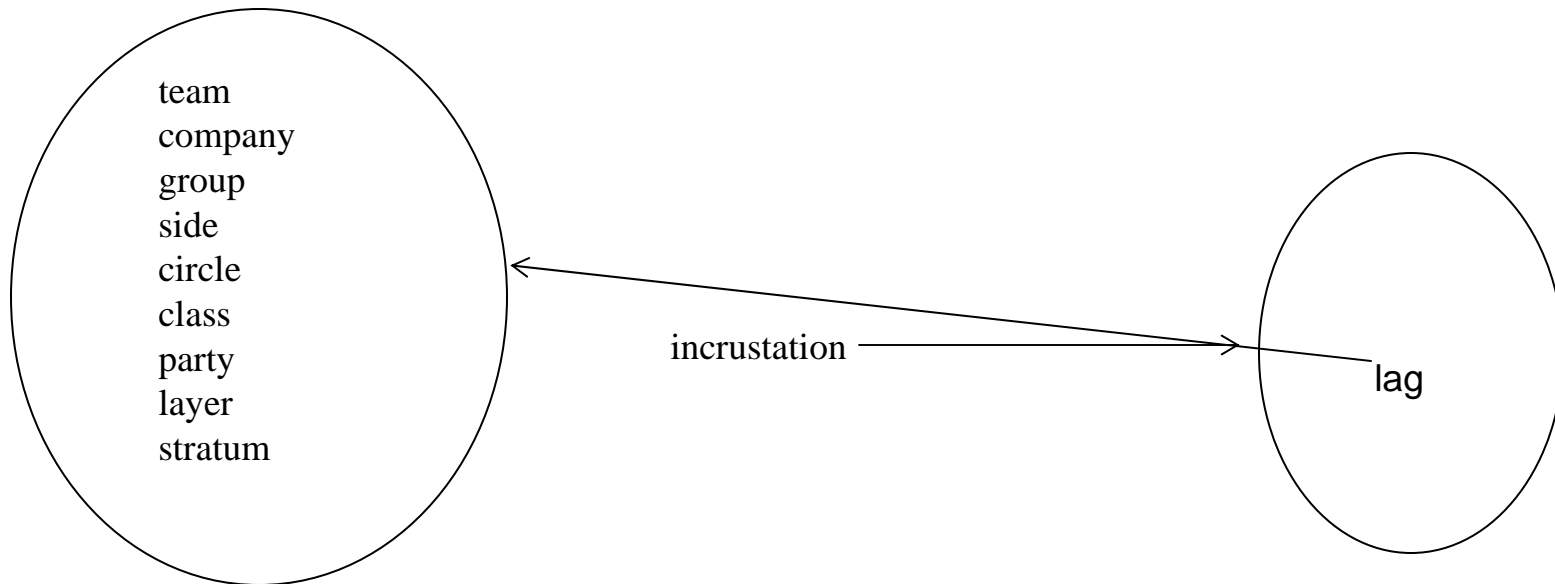
Conditions::

1. At least 3 and at least 1/4 of these words point to *solid*
2. At least one of the words pointing to *solid*, is pointed to by at least 2 words in the original *t*-image.

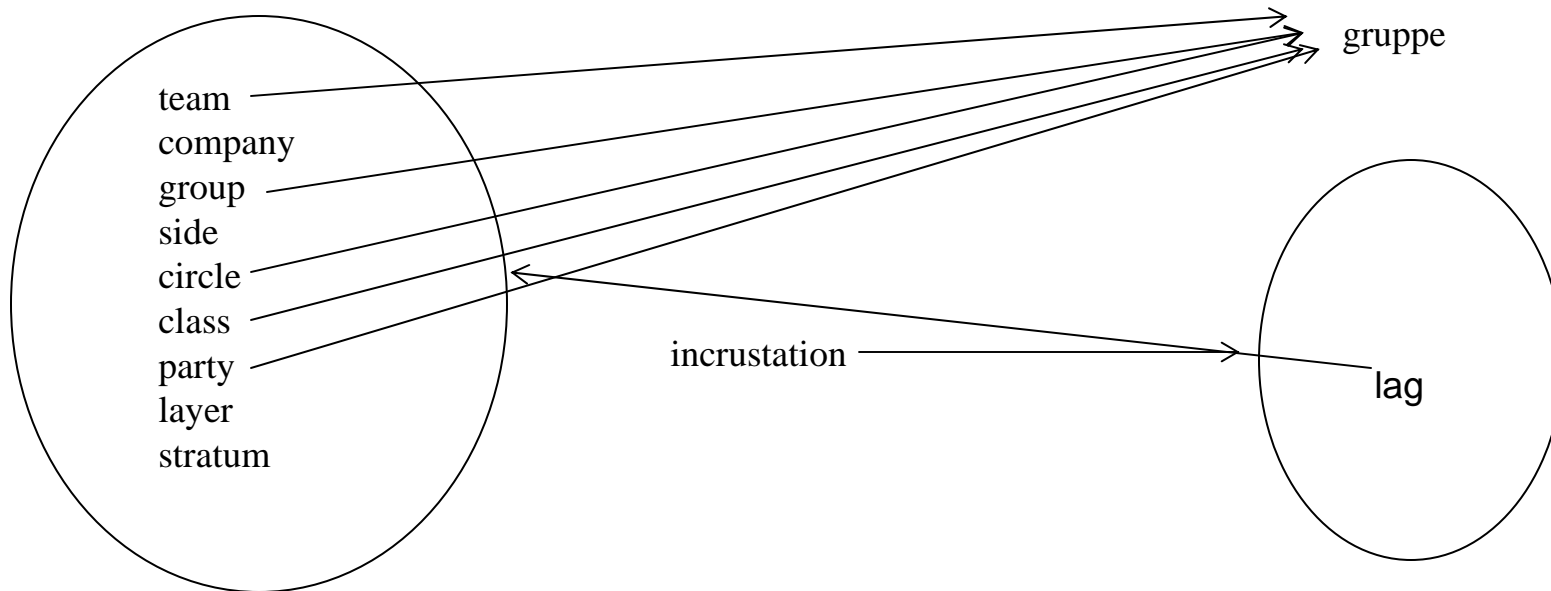
The reason why at least 2 words in the first t -image must take part in pointing out the new word:



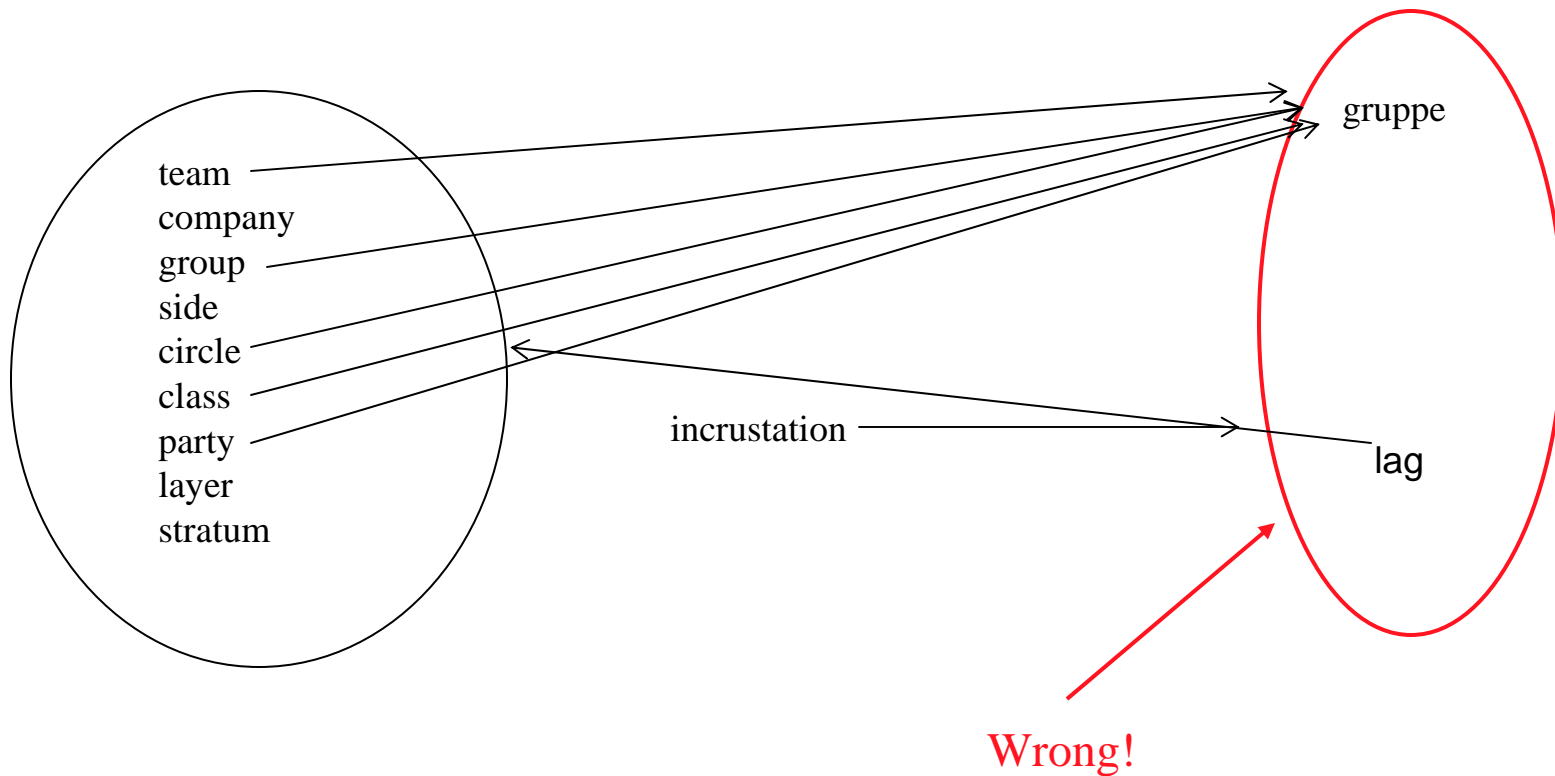
The reason why at least 2 words in the first t -image must take part in pointing out the new word:



The reason why at least 2 words in the first t -image must take part in pointing out the new word:



The reason why at least 2 words in the first t -image must take part in pointing out the new word:



Some heuristically extended *t*-images:

fantastisk: {amazing, brilliant, enormous, extraordinary, fantastic, glorious, great, magnificent, marvellous, remarkable, spectacular, splendid, wonderful} **Added:** {exceptional, startling, surprising, unusual}

firma: {business, company, establishment, firmN, place} **Added:** {factory}

full: {absolute, alive, bright, busy, complete, covered, crammed, crowded, drunk, full, full-time, great, grown, heavy, high, immediate, laden, loaded, packed, perfect, rich, teeming, total} **Added:** {entire, unified, whole}

lysende: {bright, brilliant, festive, glittering, glowing, illuminated, intense, luminous, shining} **Added:** {shiny}

pen: {attractive, beautiful, clean, elegant, fine, formal, gentle, good, handsome, lovely, neat, nice, pleasant, pretty, soft, sweet, tidy} **Added:** {charming, cute, dishy, enchanting, fancy, first-class, graceful, picturesque, polite}

selskap: {business, circle, class, companionship, company, firmN, gathering, party, society} **Added:** {cell, community, crowd, group, team}

skikkelig: {adequate, bad, competent, decent, effective, fair, firmA, genuine, good, nice, proper, real, regular, respectable, satisfactory, smart, sure, true} **Added:** {accurate, honest, legitimate, sincere}

Extended *t*-images (cont.):

stødig: {firm, steady} **Added:** {solid, strong}

strålende: {beautiful, bright, brilliant, excellent, fine, gleaming, glistening, glorious, glowing, great, magnificent, radiant, sparkling, splendid, sumptuous, warm} **Added:** {fantastic, glittering, shining, shiny}

vedvarende: {abiding, constant, continuous, persistent, steady, unresolved} **Added:** {perpetual}

vennlig: {amiable, benign, bland, cheerful, friendly, gentle, kind, kindly, nice, pleasant, pleasing, sympathetic} **Added:** {good-humoured, good-natured, polite, smiling}

viktig: {all-important, basic, big, central, considerable, critical, crucial, essential, great, helpful, high, important, key, large, lively, major, massive, momentous, notable, pompous, significant, urgent, vital, worthy} **Added:** {decisive, outstanding, primary, supreme}

virkelig: {actual, biological, confident, even, genuine, habitual, physical, proper, real, regular, right, serious, true} **Added:** {accurate, authentic, truthful}

Assumptions

1. Semantically closely related words tend to have strongly overlapping sets of translations.

Assumptions

1. Semantically closely related words tend to have strongly overlapping sets of translations.
2. Words with wide meanings tend to have a higher number of translations than words with narrow meanings.

Assumptions

1. Semantically closely related words tend to have strongly overlapping sets of translations.
2. Words with wide meanings tend to have a higher number of translations than words with narrow meanings.
3. If a word *a* is a hyponym of a word *b* (such as *tasty* of *good*, for example), then the possible translations of *a* will probably be a subset of the possible translations of *b*.

Assumptions

1. Semantically closely related words tend to have strongly overlapping sets of translations.
2. Words with wide meanings tend to have a higher number of translations than words with narrow meanings.
3. If a word *a* is a hyponym of a word *b* (such as *tasty* of *good*, for example), then the possible translations of *a* will probably be a subset of the possible translations of *b*.
4. Contrastive ambiguity, i.e., ambiguity between two unrelated senses of a word, such as the two senses of the English noun *bank* ('money institution' and 'riverside'), tends to be a historically accidental and idiosyncratic property of individual words. Hence we don't expect to find instances of the same contrastive ambiguity replicated by other words in the language or by words in other languages.

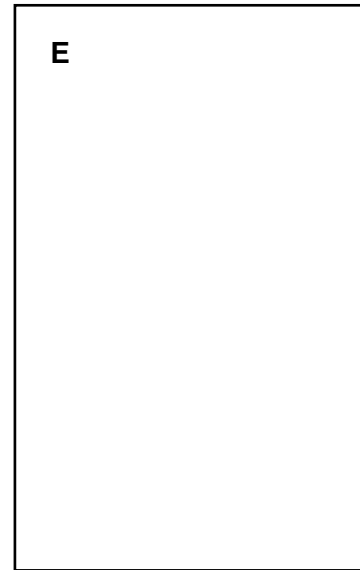
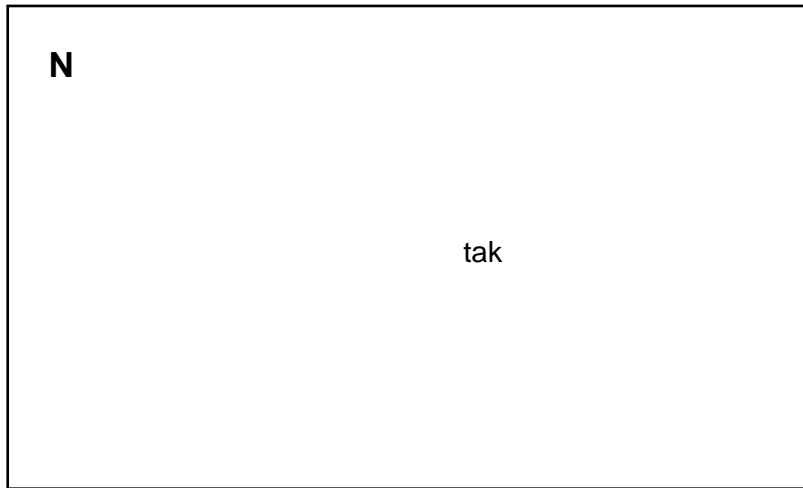
Assumptions

1. Semantically closely related words tend to have strongly overlapping sets of translations.
2. Words with wide meanings tend to have a higher number of translations than words with narrow meanings.
3. If a word *a* is a hyponym of a word *b* (such as *tasty* of *good*, for example), then the possible translations of *a* will probably be a subset of the possible translations of *b*.
4. Contrastive ambiguity, i.e., ambiguity between two unrelated senses of a word, such as the two senses of the English noun *bank* ('money institution' and 'riverside'), tends to be a historically accidental and idiosyncratic property of individual words. Hence we don't expect to find instances of the same contrastive ambiguity replicated by other words in the language or by words in other languages. (More precisely, we should talk about ambiguous *phonological/graphic* words here, since such ambiguity is normally analysed as homonymy and hence as involving two lemmas.)

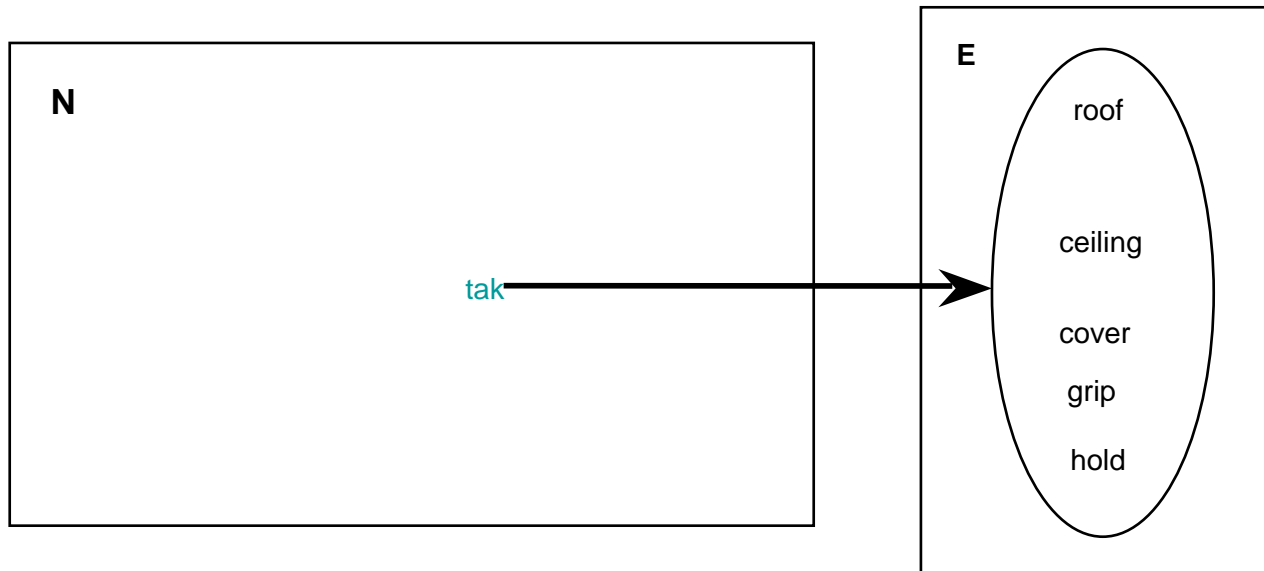
Assumptions

1. Semantically closely related words tend to have strongly overlapping sets of translations.
2. Words with wide meanings tend to have a higher number of translations than words with narrow meanings.
3. If a word *a* is a hyponym of a word *b* (such as *tasty* of *good*, for example), then the possible translations of *a* will probably be a subset of the possible translations of *b*.
4. Contrastive ambiguity, i.e., ambiguity between two unrelated senses of a word, such as the two senses of the English noun *bank* ('money institution' and 'riverside'), tends to be a historically accidental and idiosyncratic property of individual words. Hence we don't expect to find instances of the same contrastive ambiguity replicated by other words in the language or by words in other languages. (More precisely, we should talk about ambiguous *phonological/graphic* words here, since such ambiguity is normally analysed as homonymy and hence as involving two lemmas.)
5. Words with unrelated meanings will not share translations into another language, except in cases where the shared word is contrastively ambiguous between the two unrelated meanings. By assumption 4 there should then be at most one such shared word.

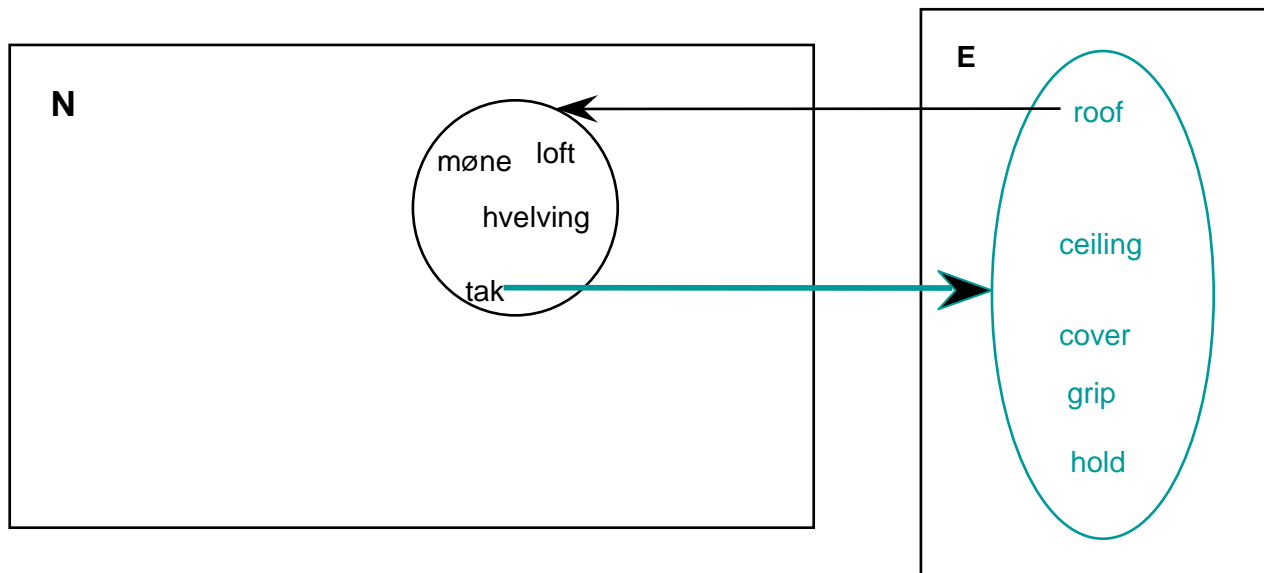
First example: 'tak'



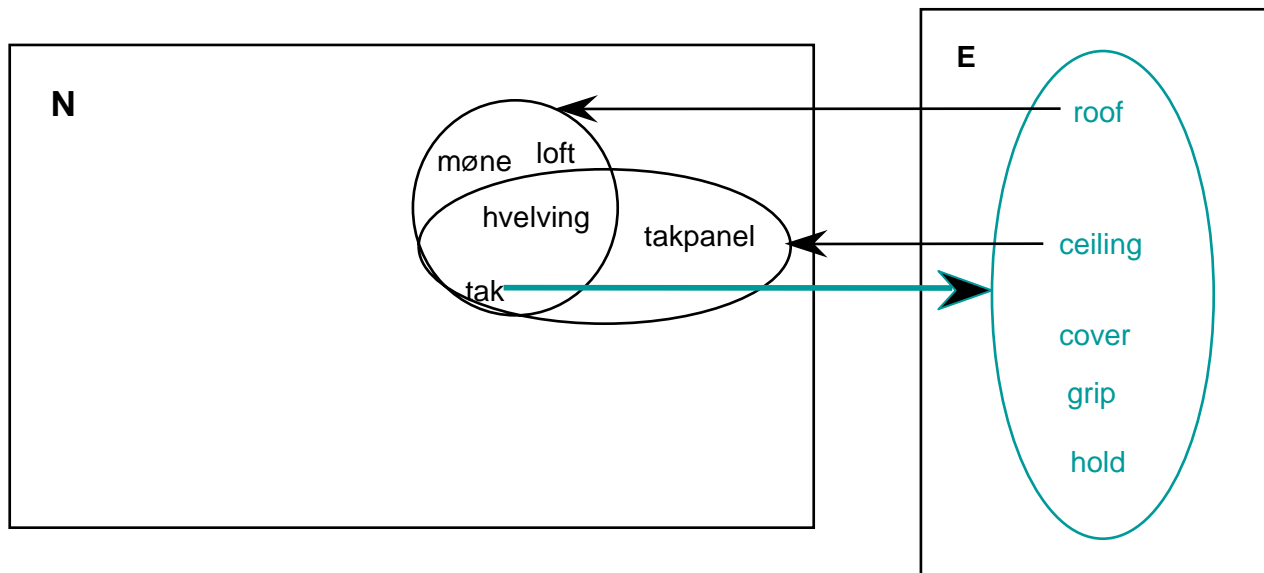
First t -image



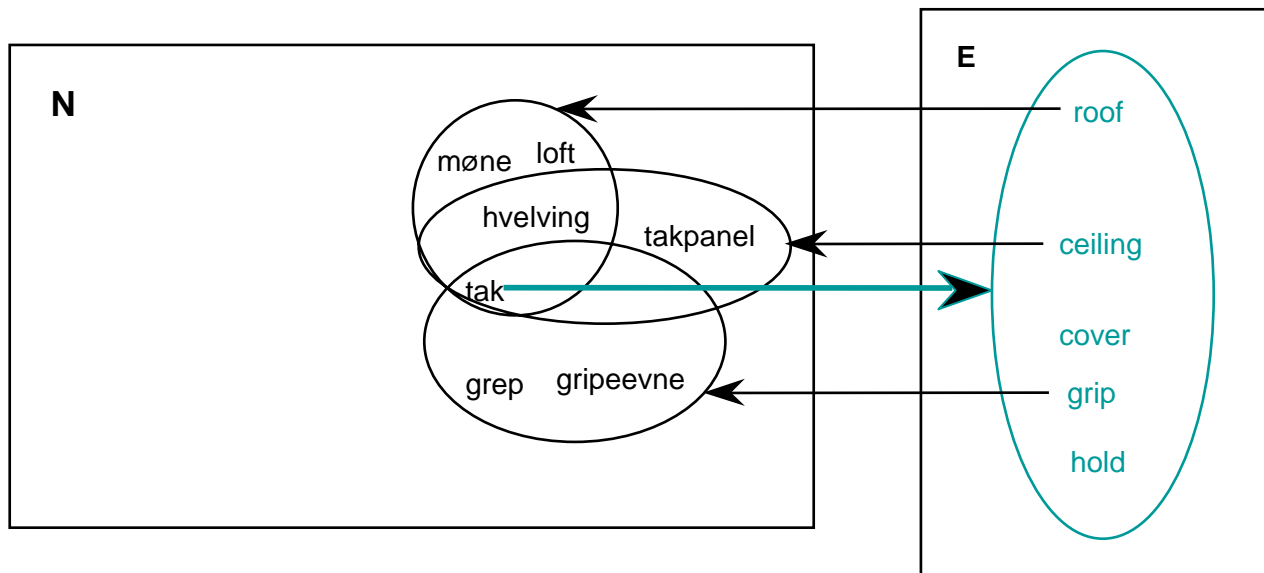
Inverse t -image



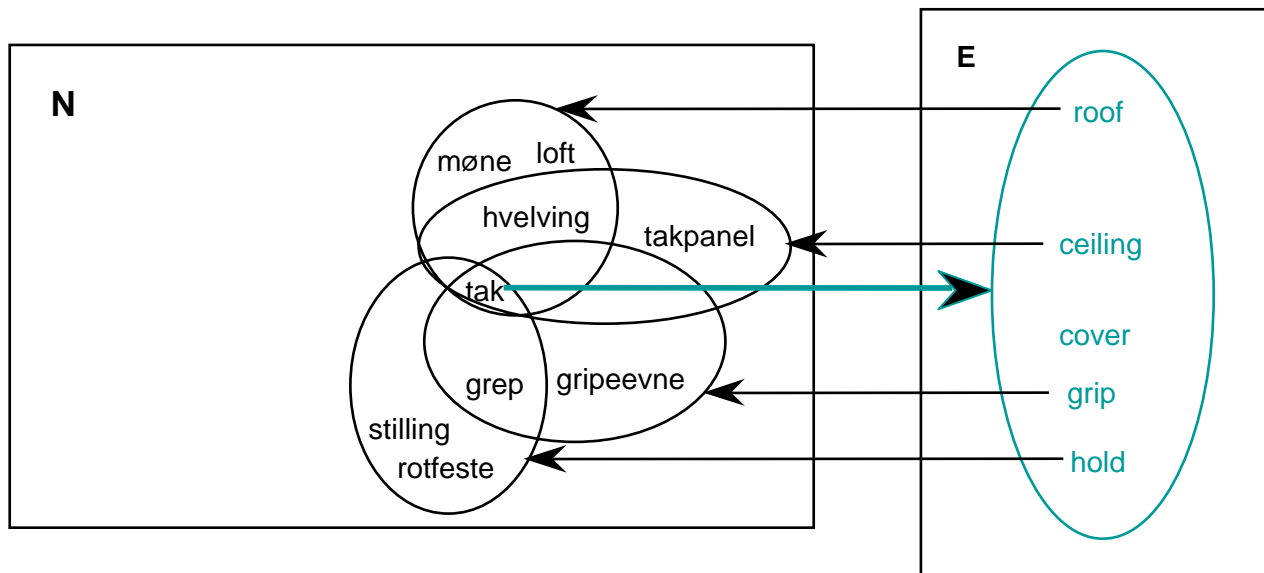
Inverse t -image



Inverse t -image



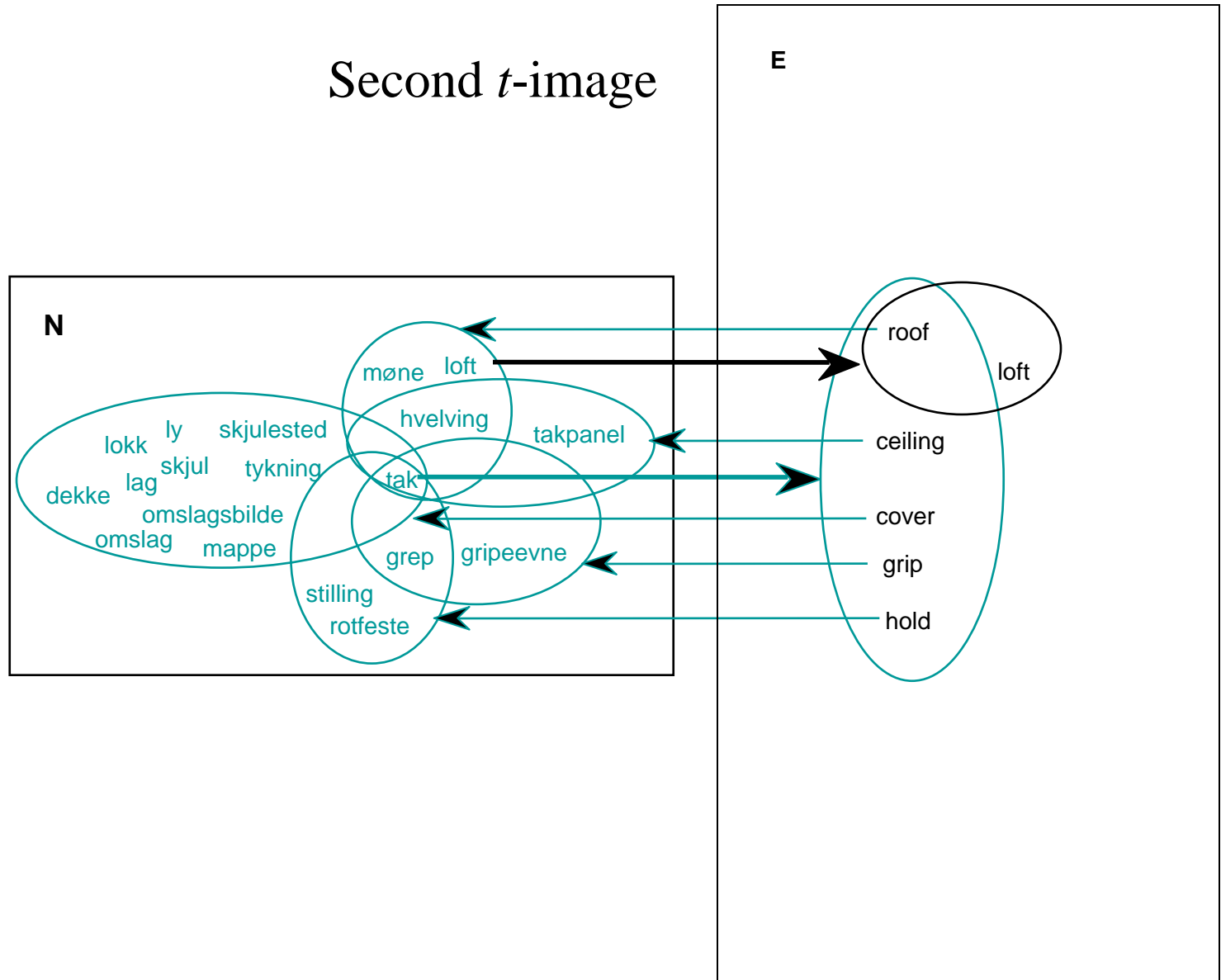
Inverse t -image



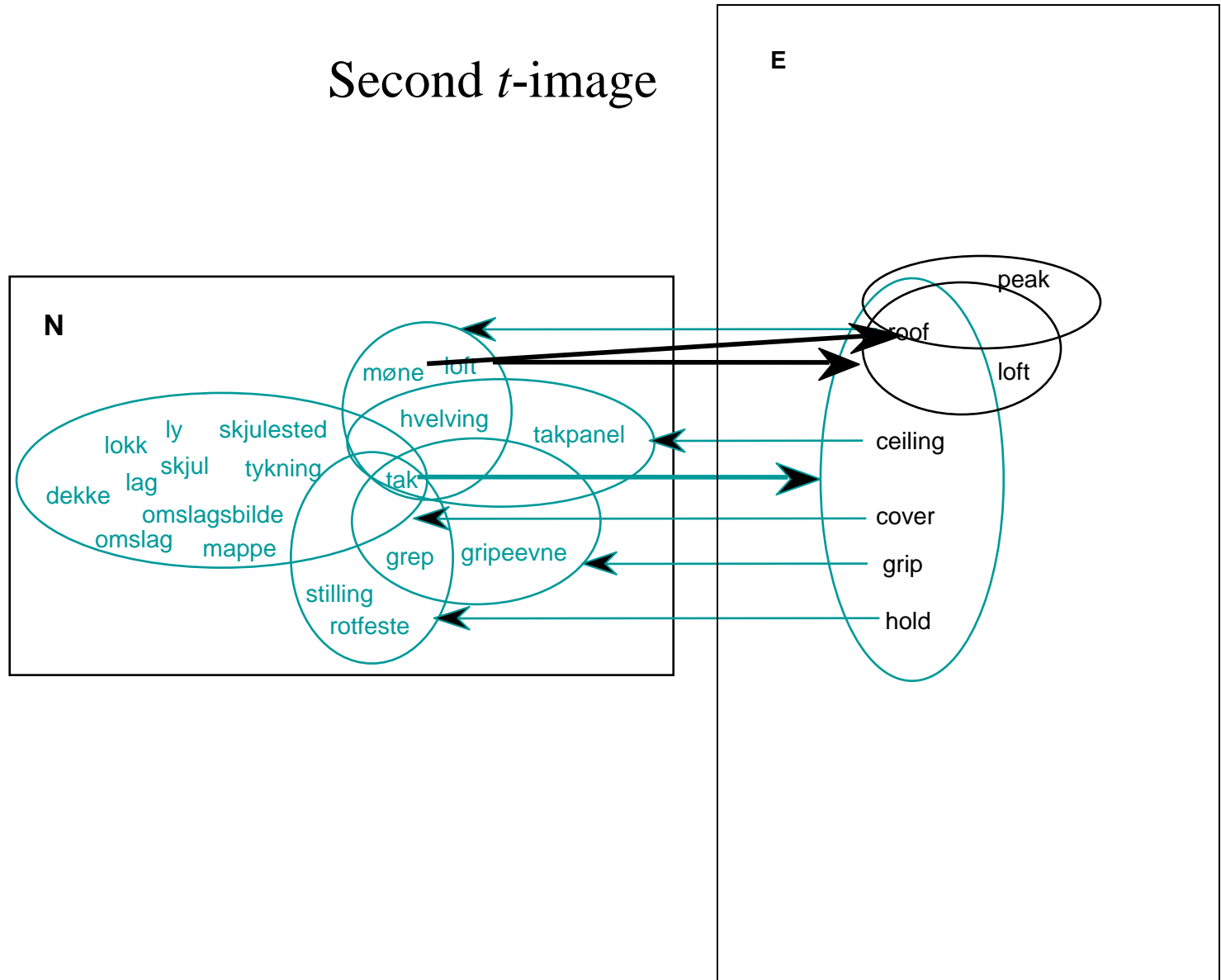
Inverse t -image



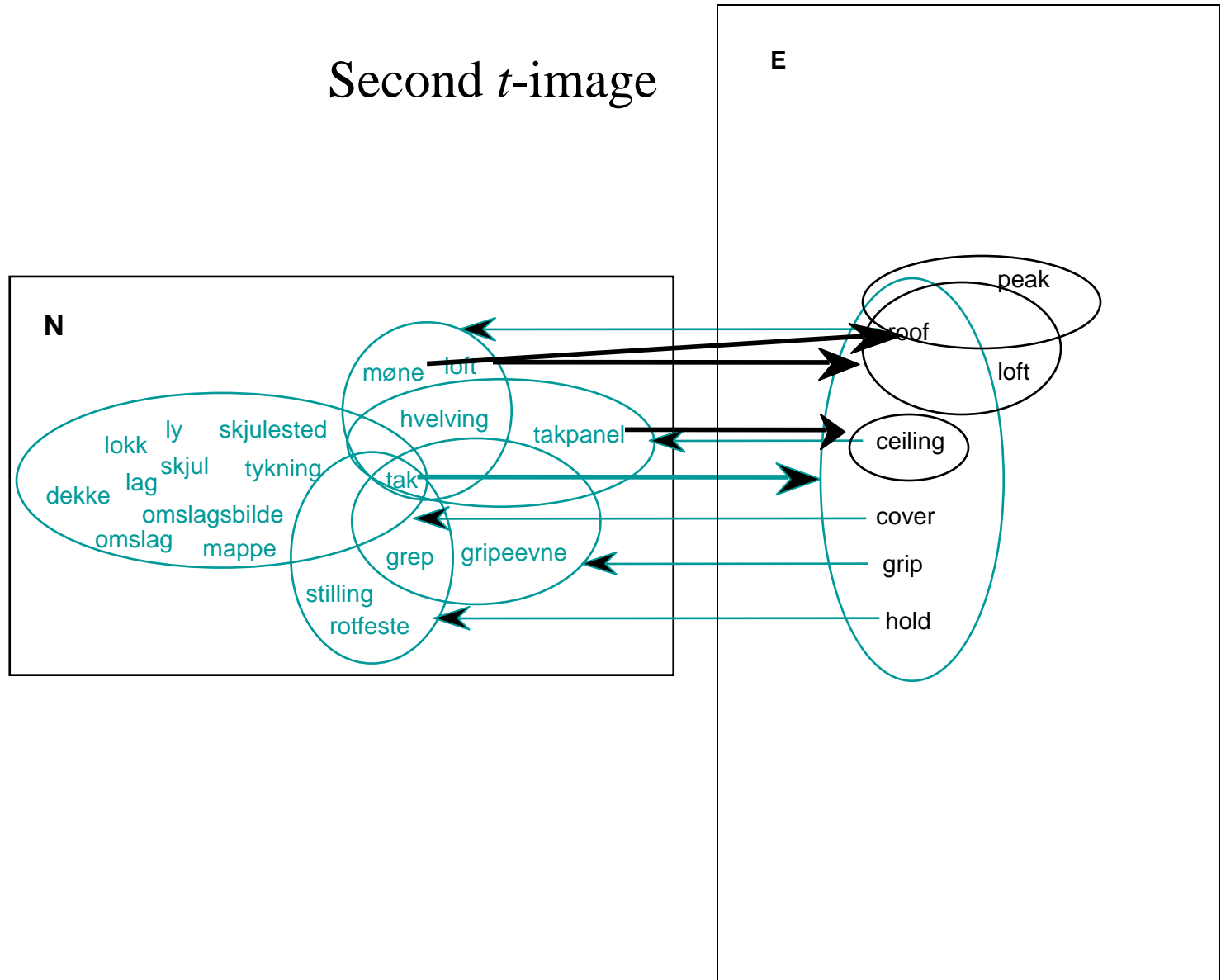
Second *t*-image



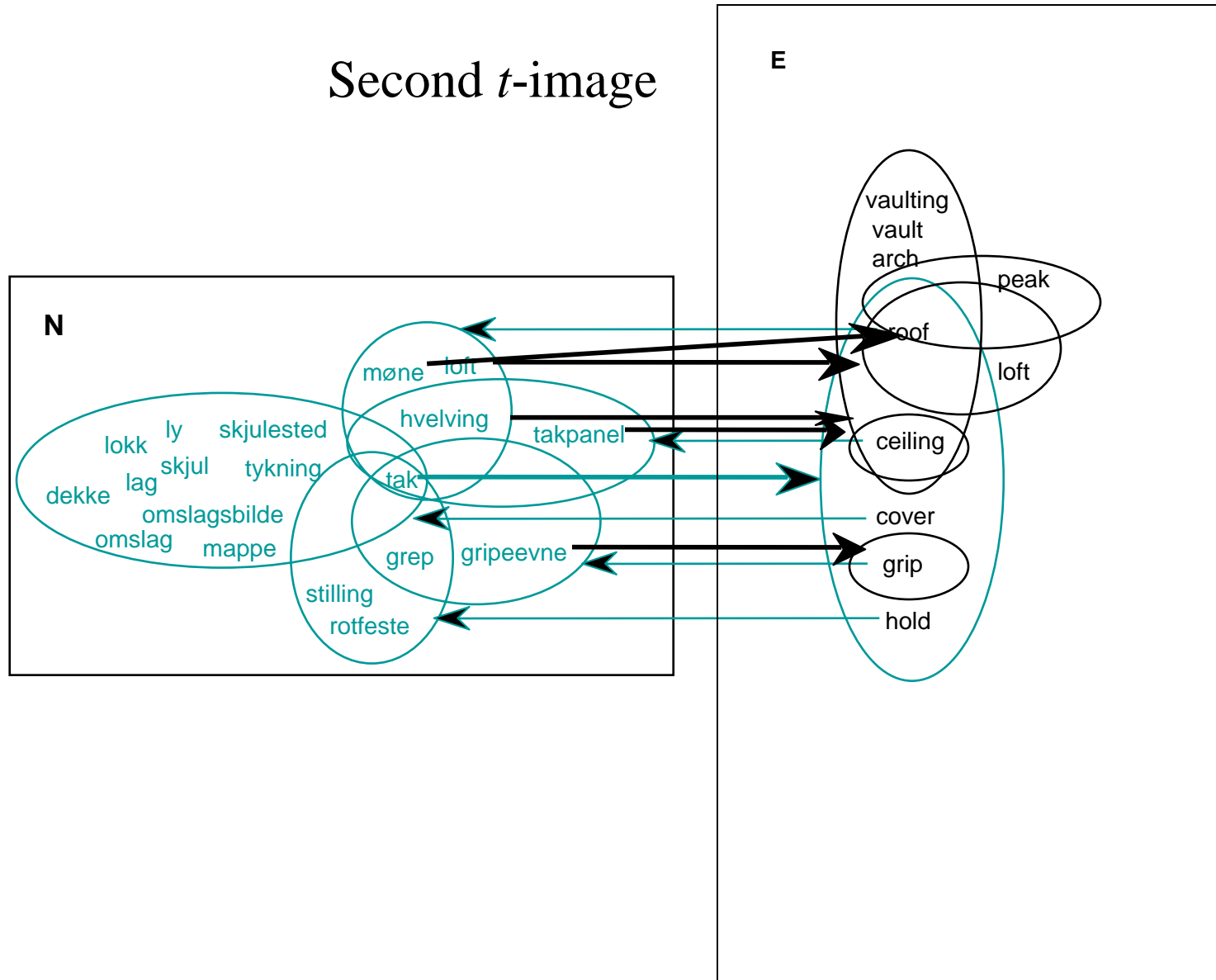
Second *t*-image



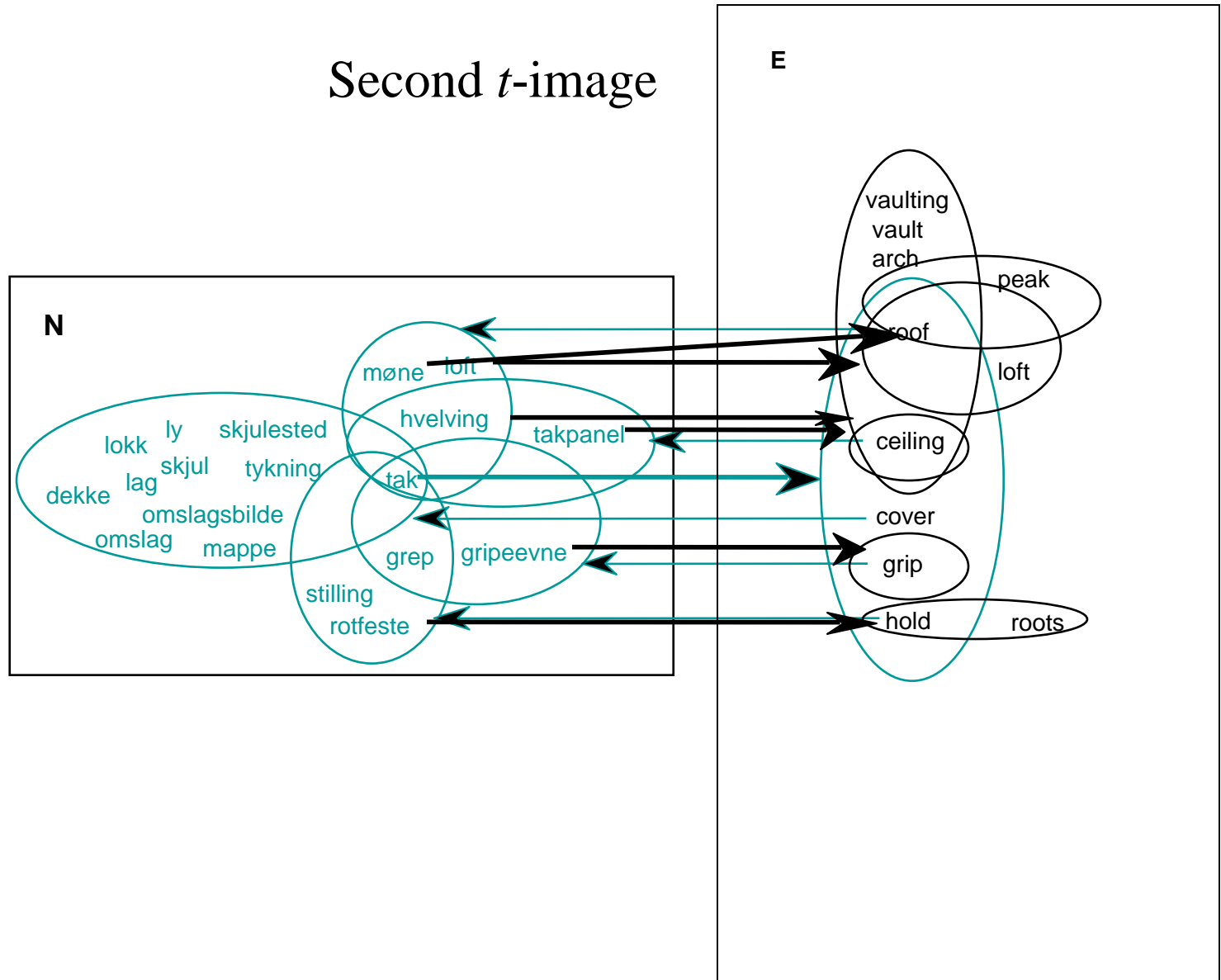
Second *t*-image



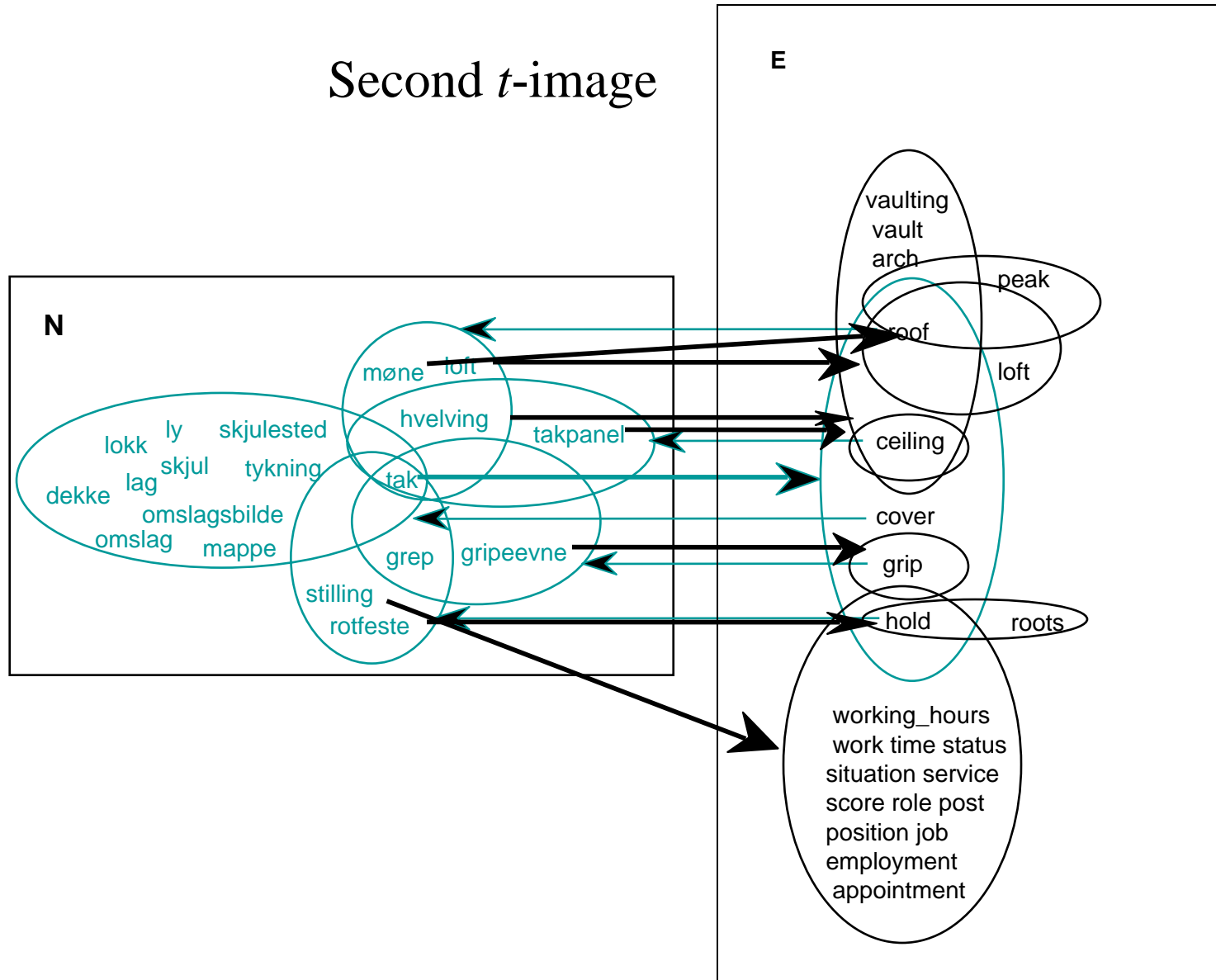
Second *t*-image



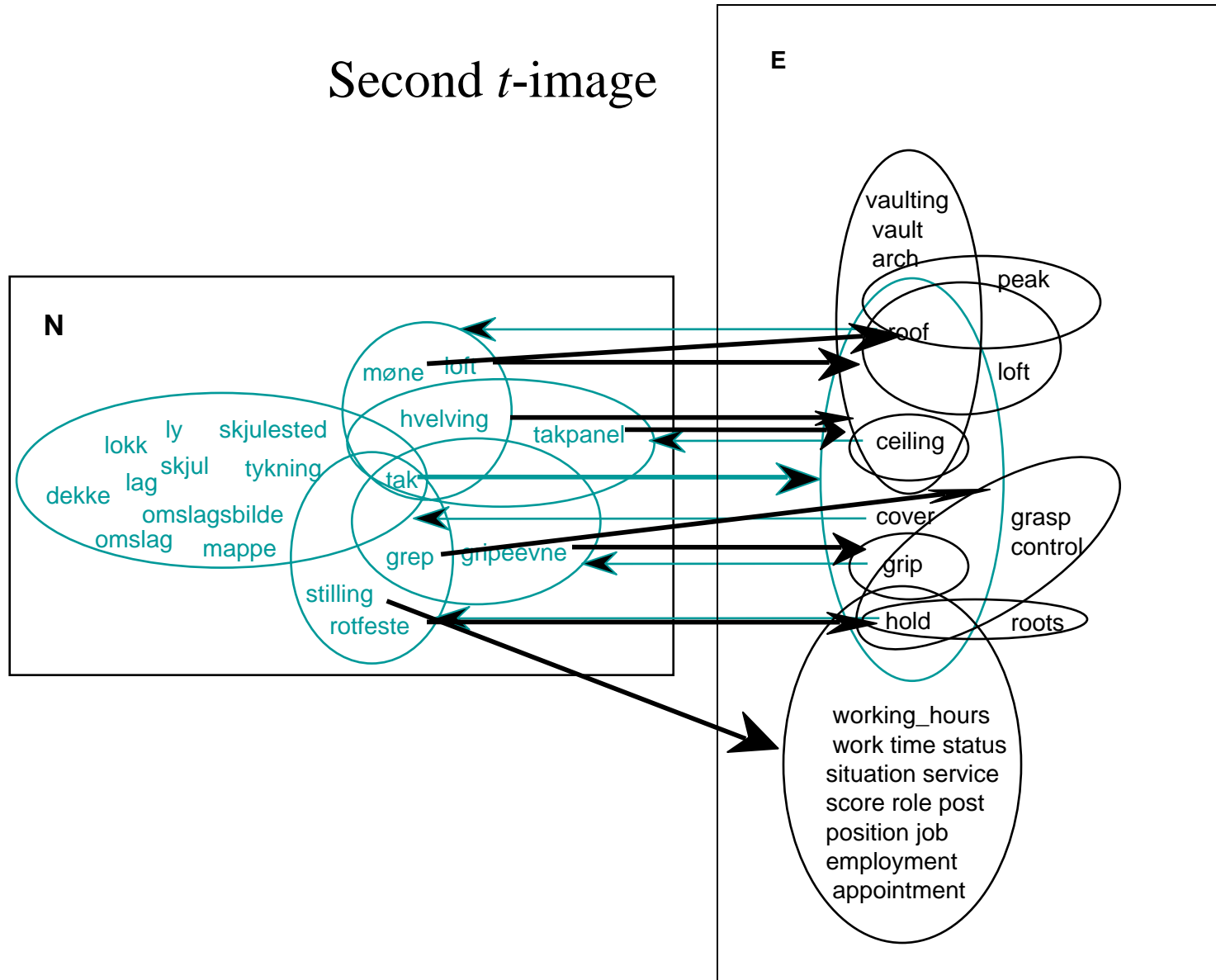
Second *t*-image



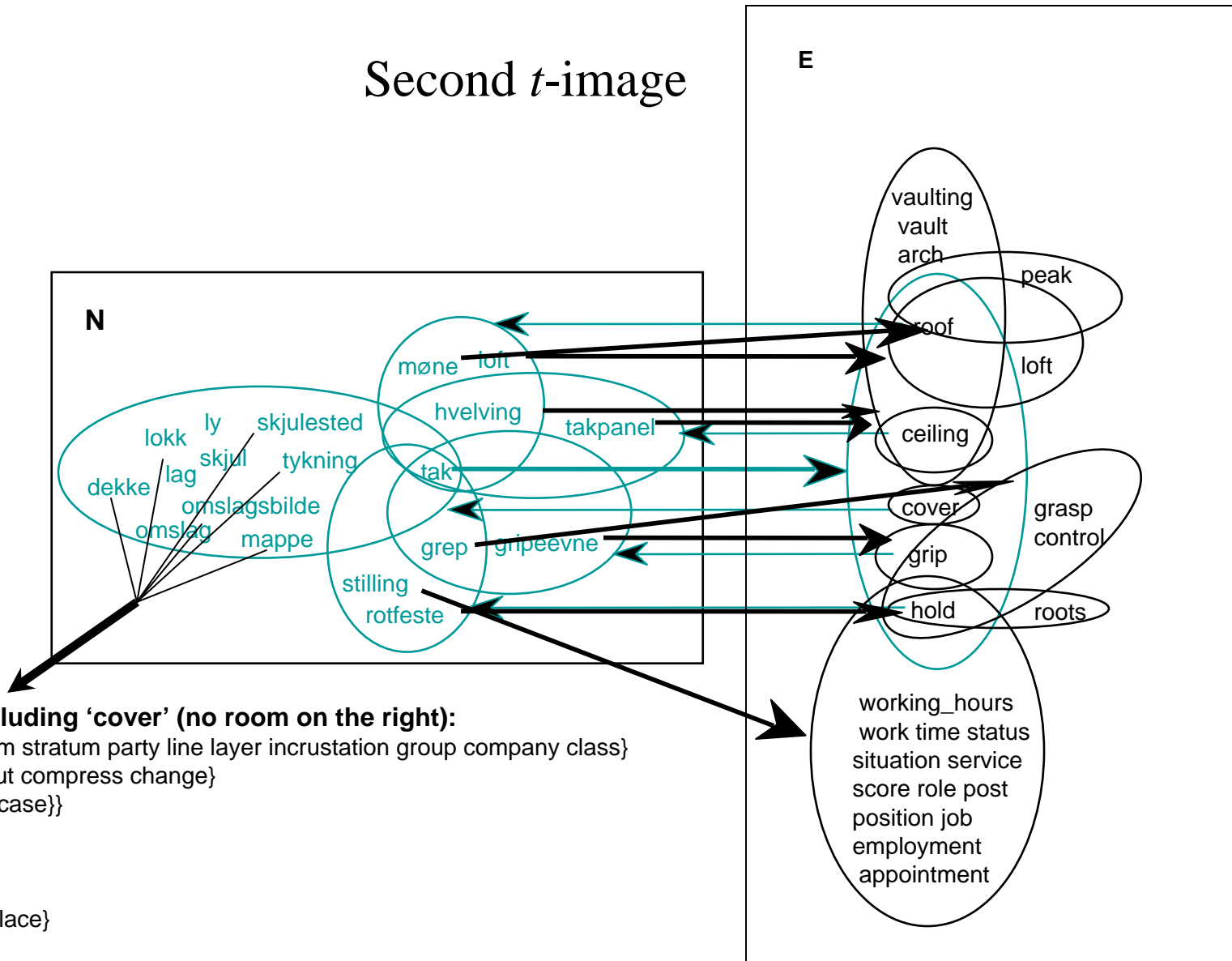
Second *t*-image



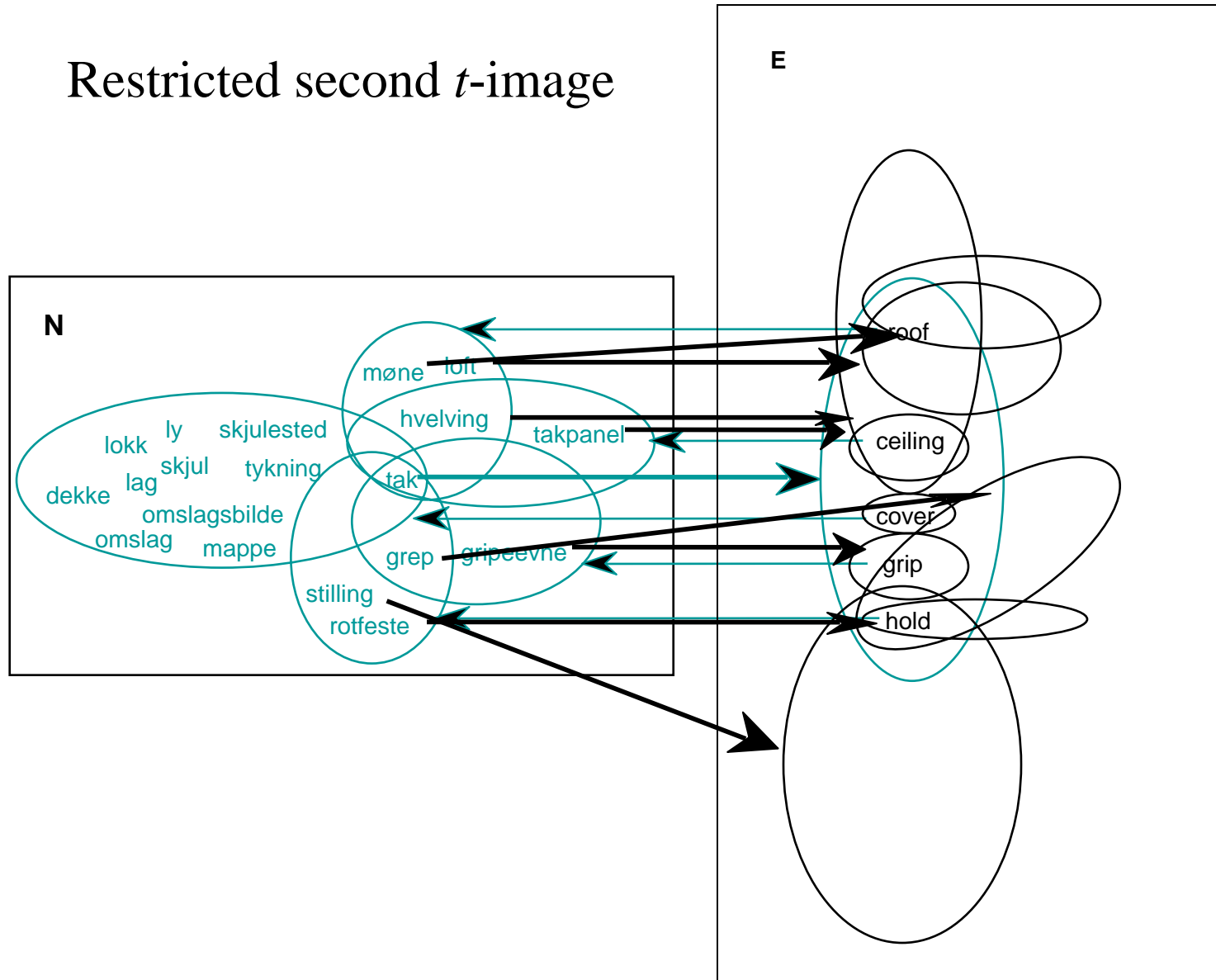
Second *t*-image



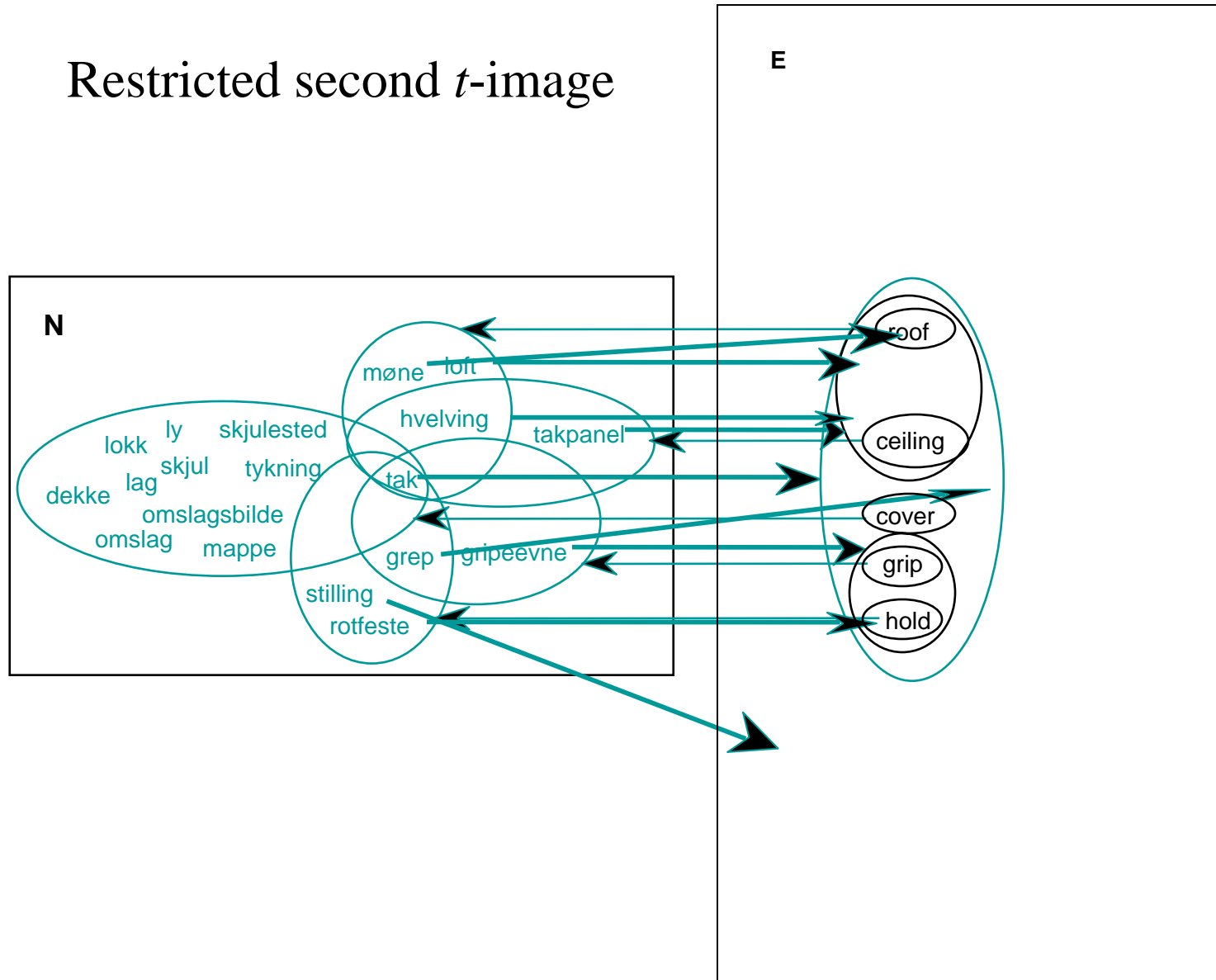
Second *t*-image



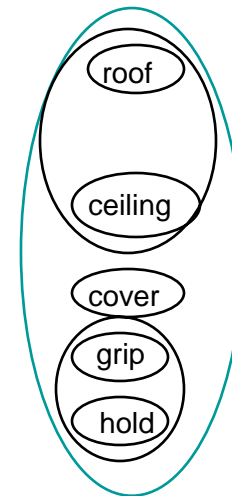
Restricted second t -image



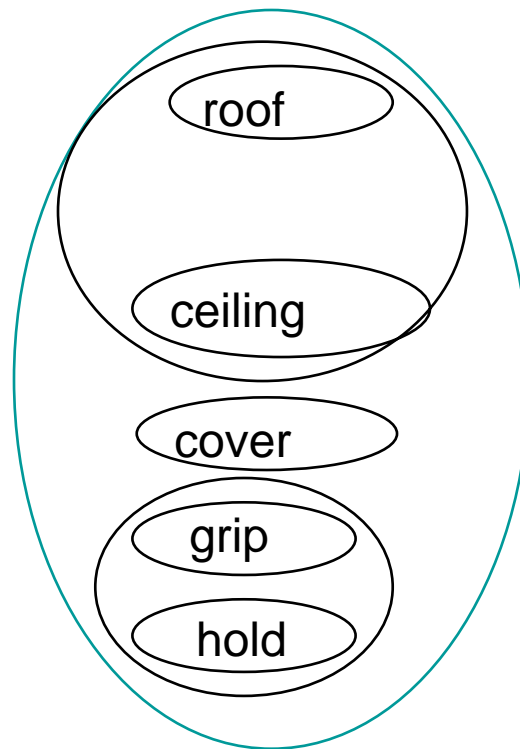
Restricted second *t*-image



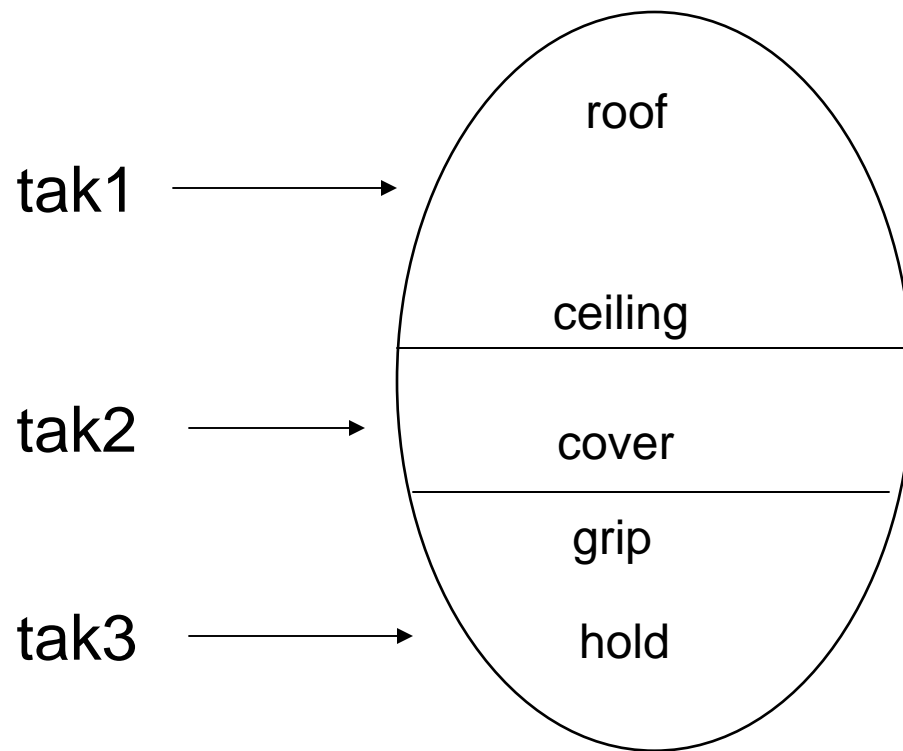
Three «sense groups» for ‘tak’
(groups of intersecting sets):



Three «sense groups» for ‘tak’
(groups of intersecting sets):



Three «sense partitions» of ‘tak’
s first *t*-image:



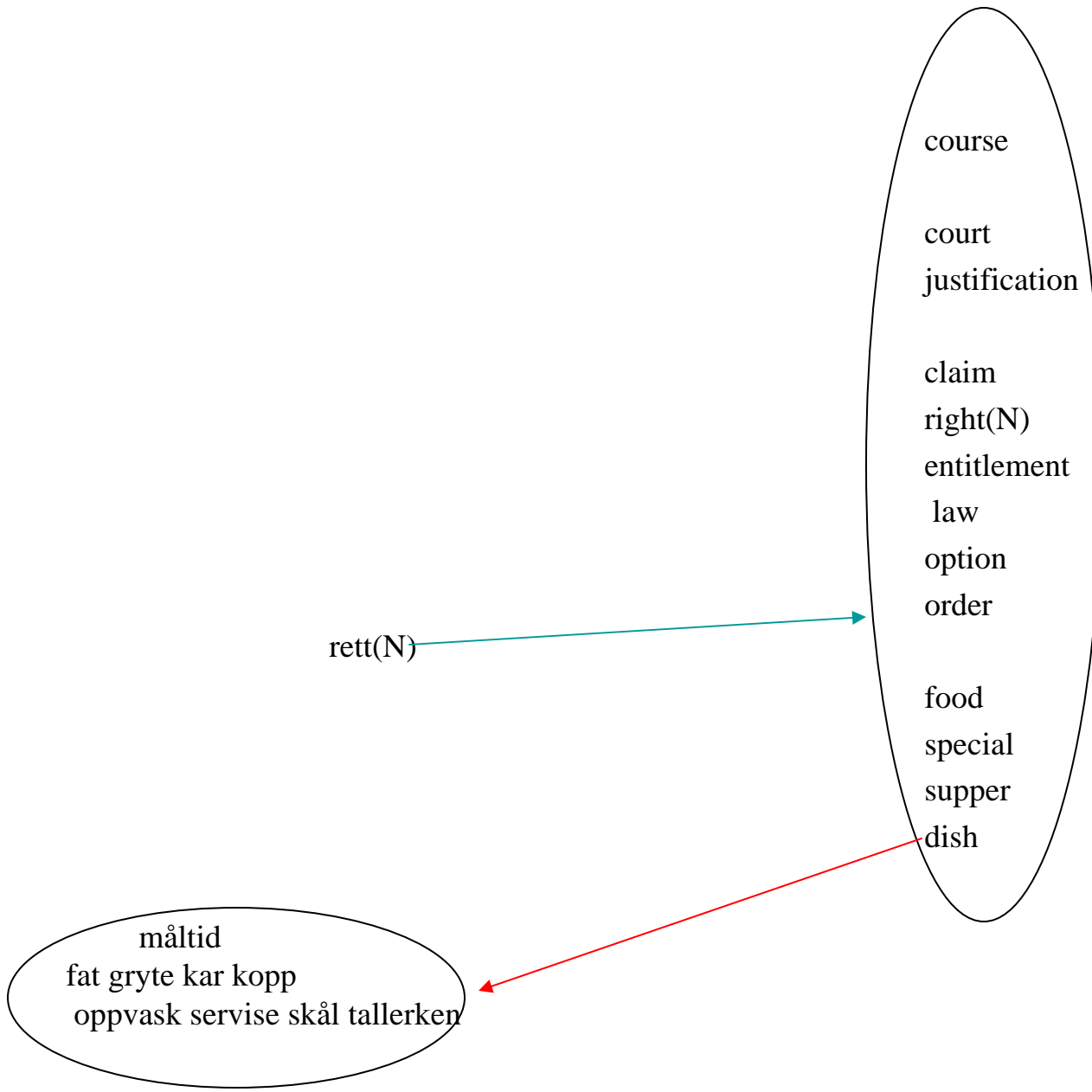
Second example: 'rett'
(Thanks to Gunn Inger Lyse)

rett(N)

rett(N)



- course
- court
- justification
- claim
- right(N)
- entitlement
- law
- option
- order
- food
- special
- supper
- dish



course

court

justification

claim

right(N)

entitlement

law

option

order

food

special

supper

dish

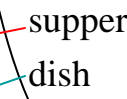
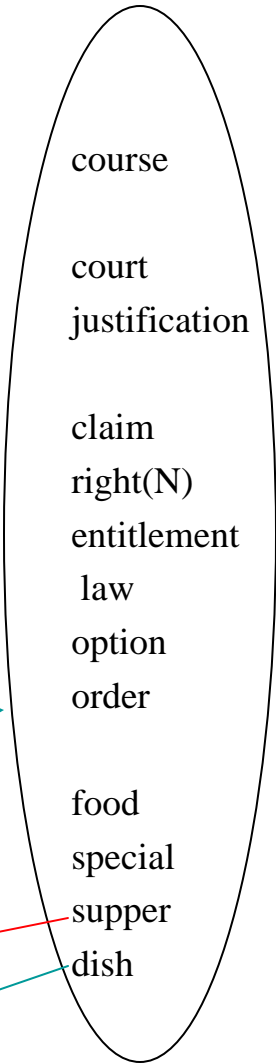
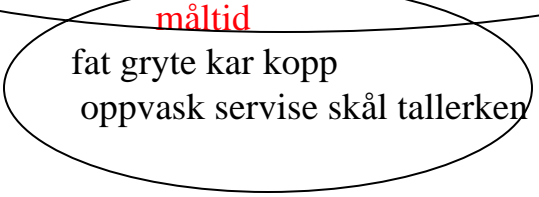
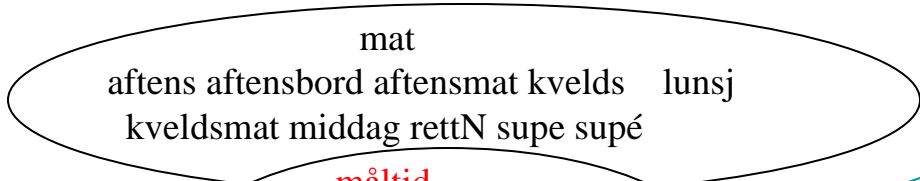
rett(N)

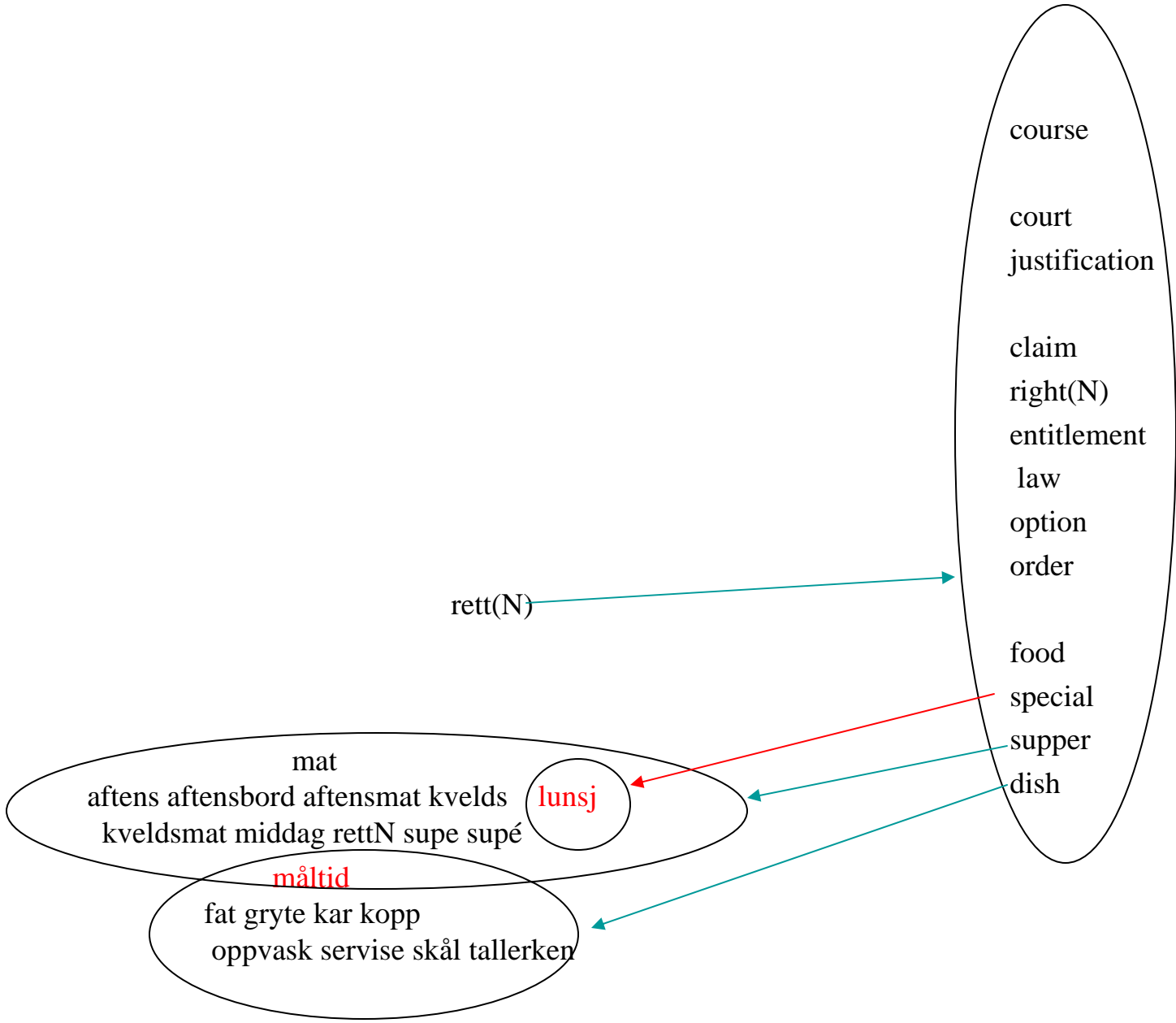
mat

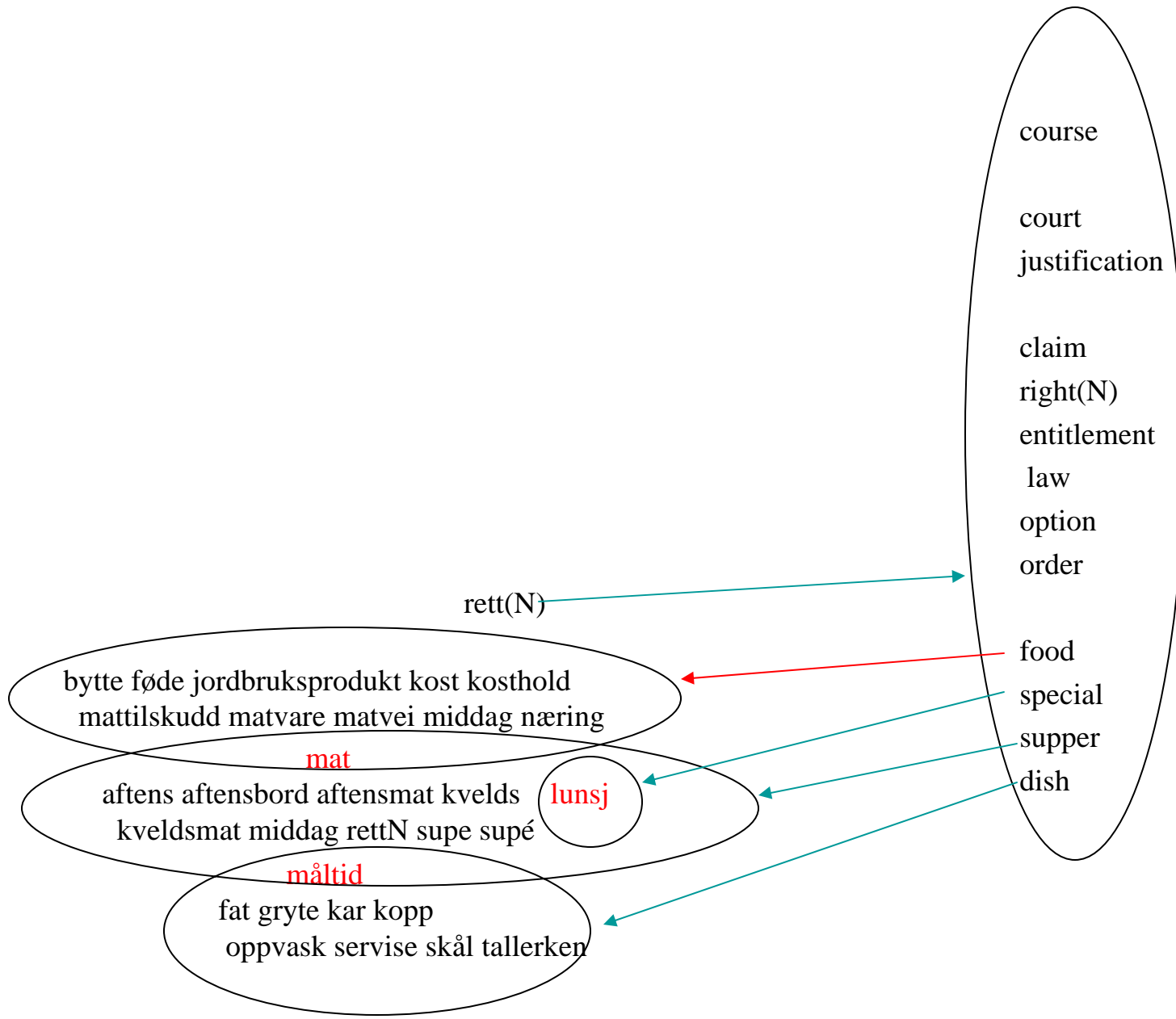
aftens aftensbord aftensmat kvelds lunsj
kveldsmat middag rettN supe supé

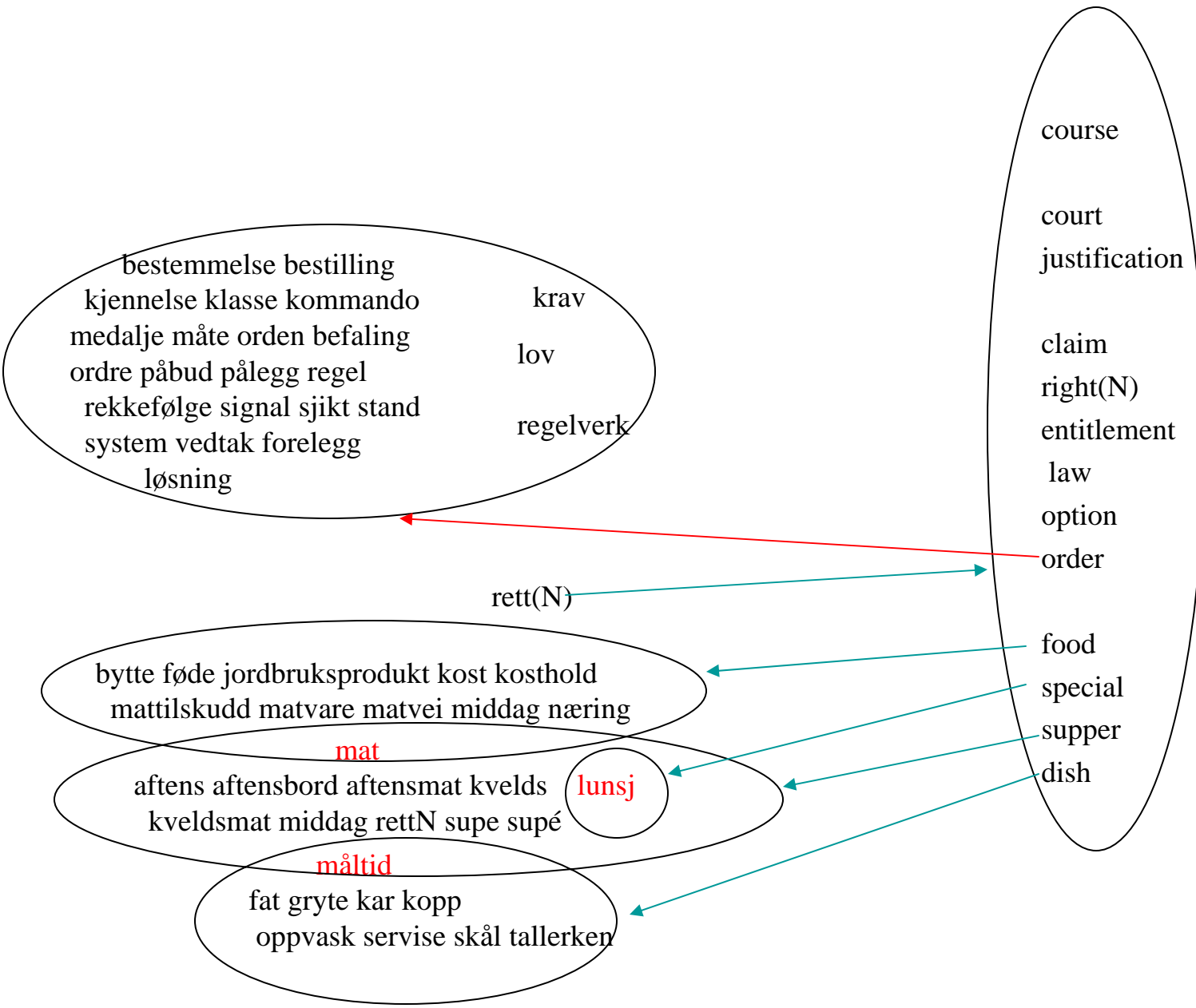
måltid

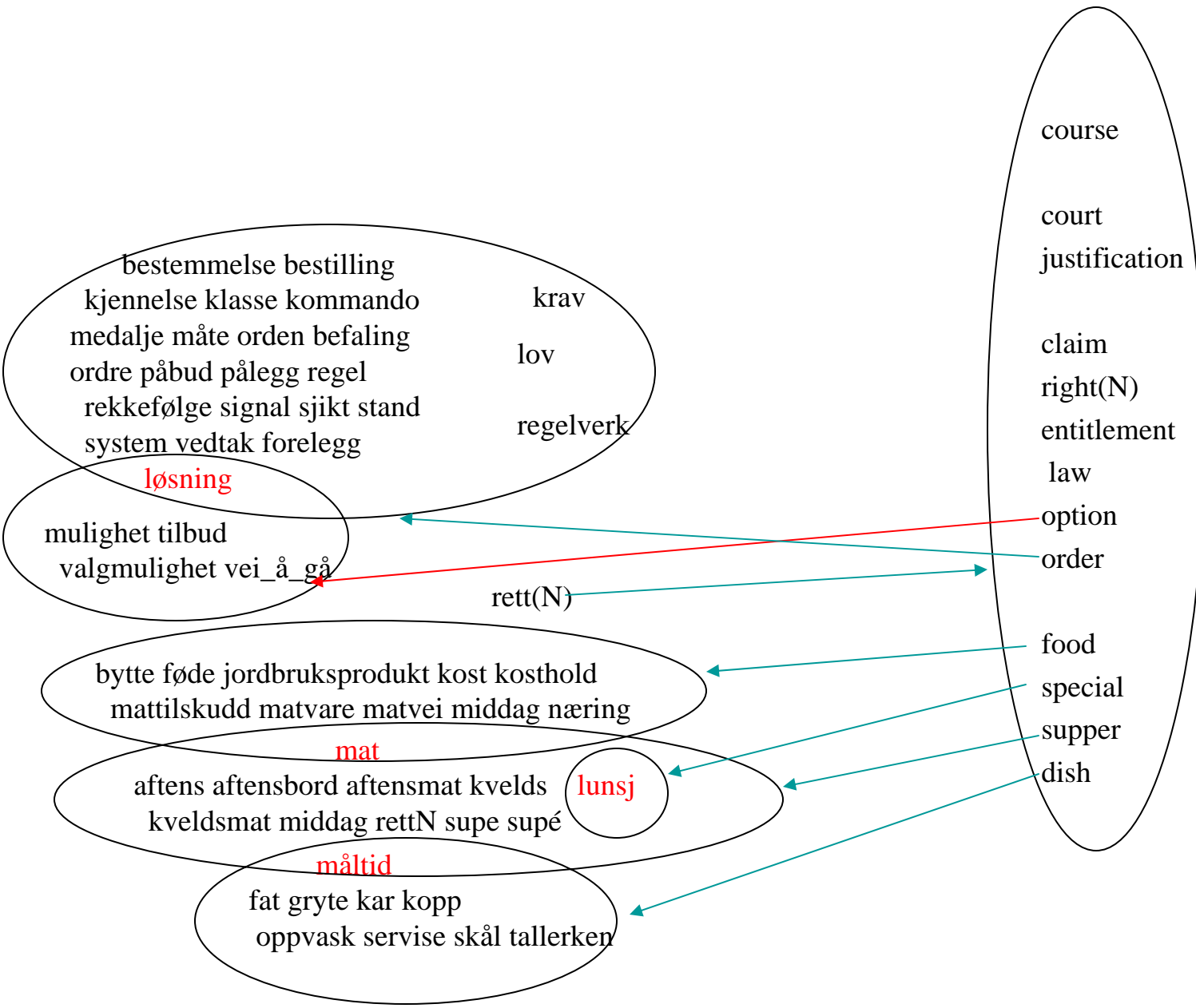
fat gryte kar kopp
oppvask servise skål tallerken

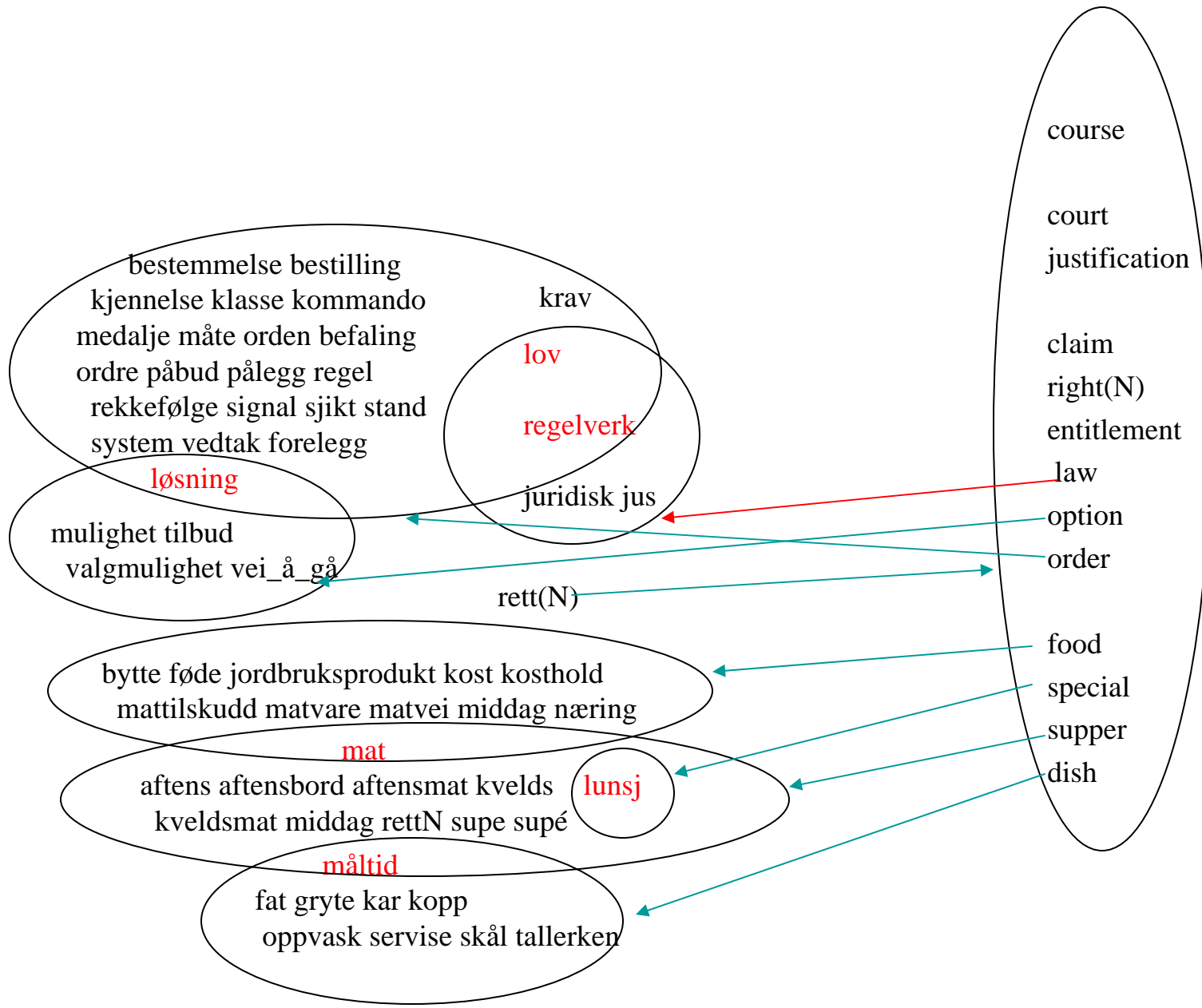


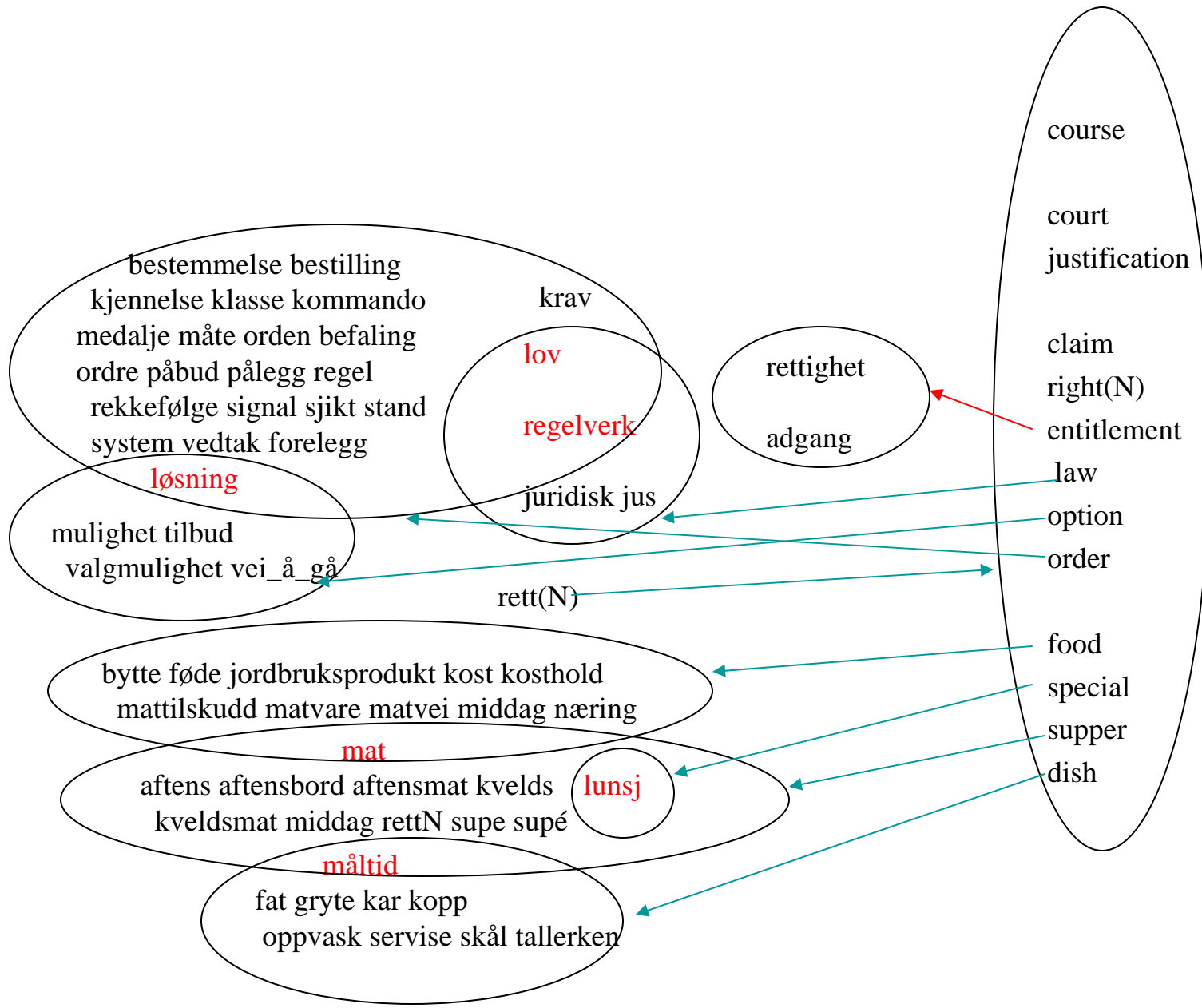


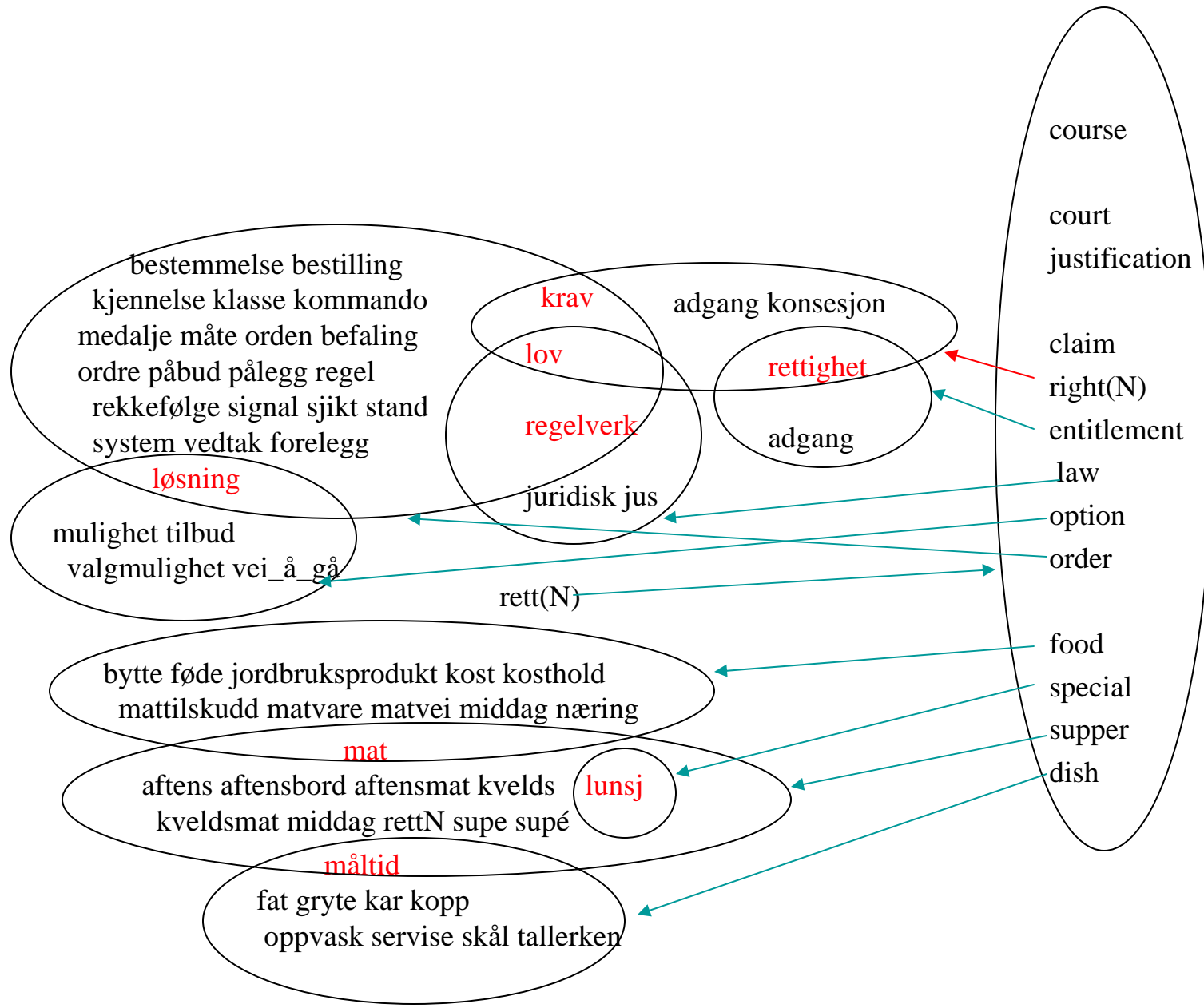


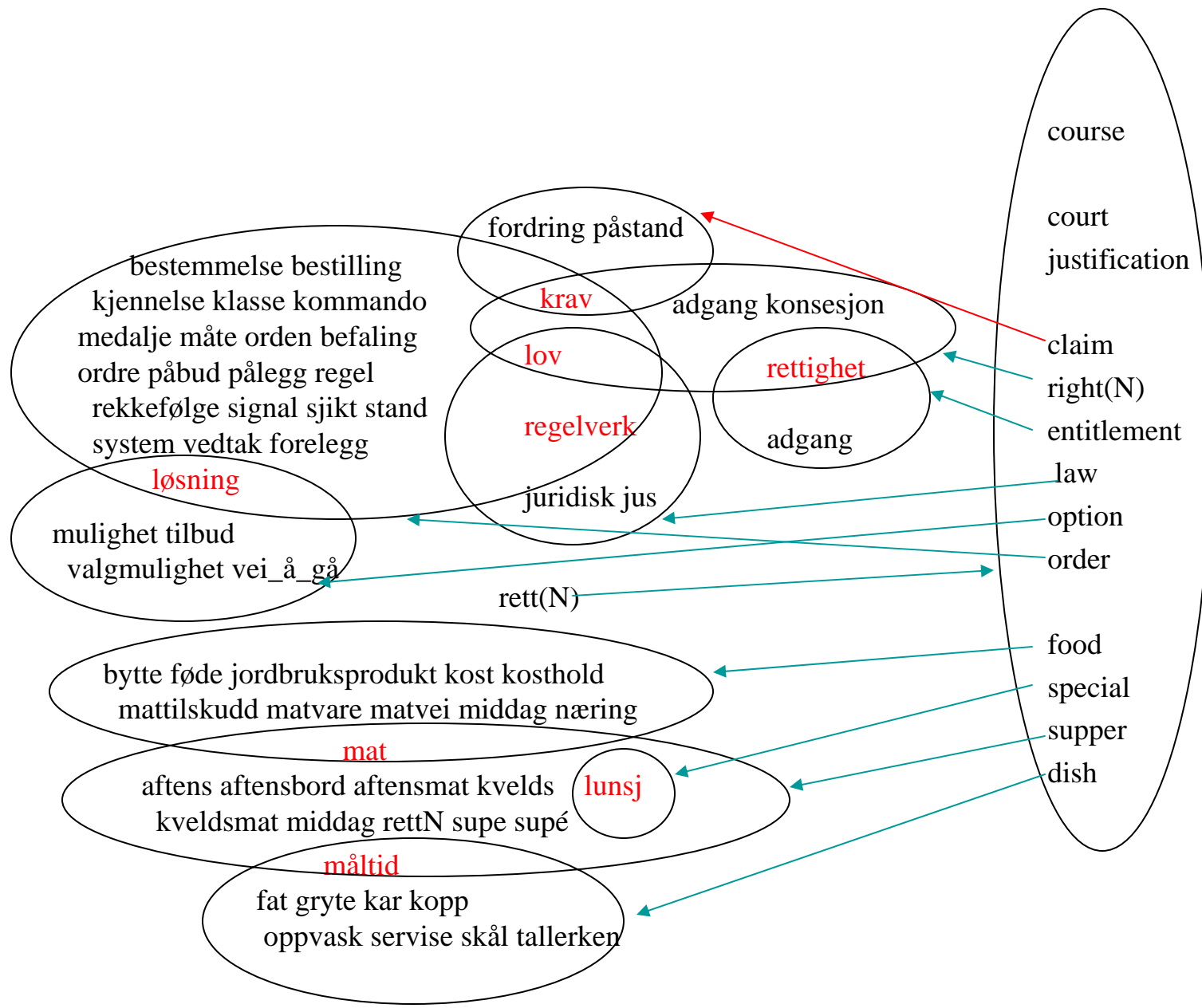


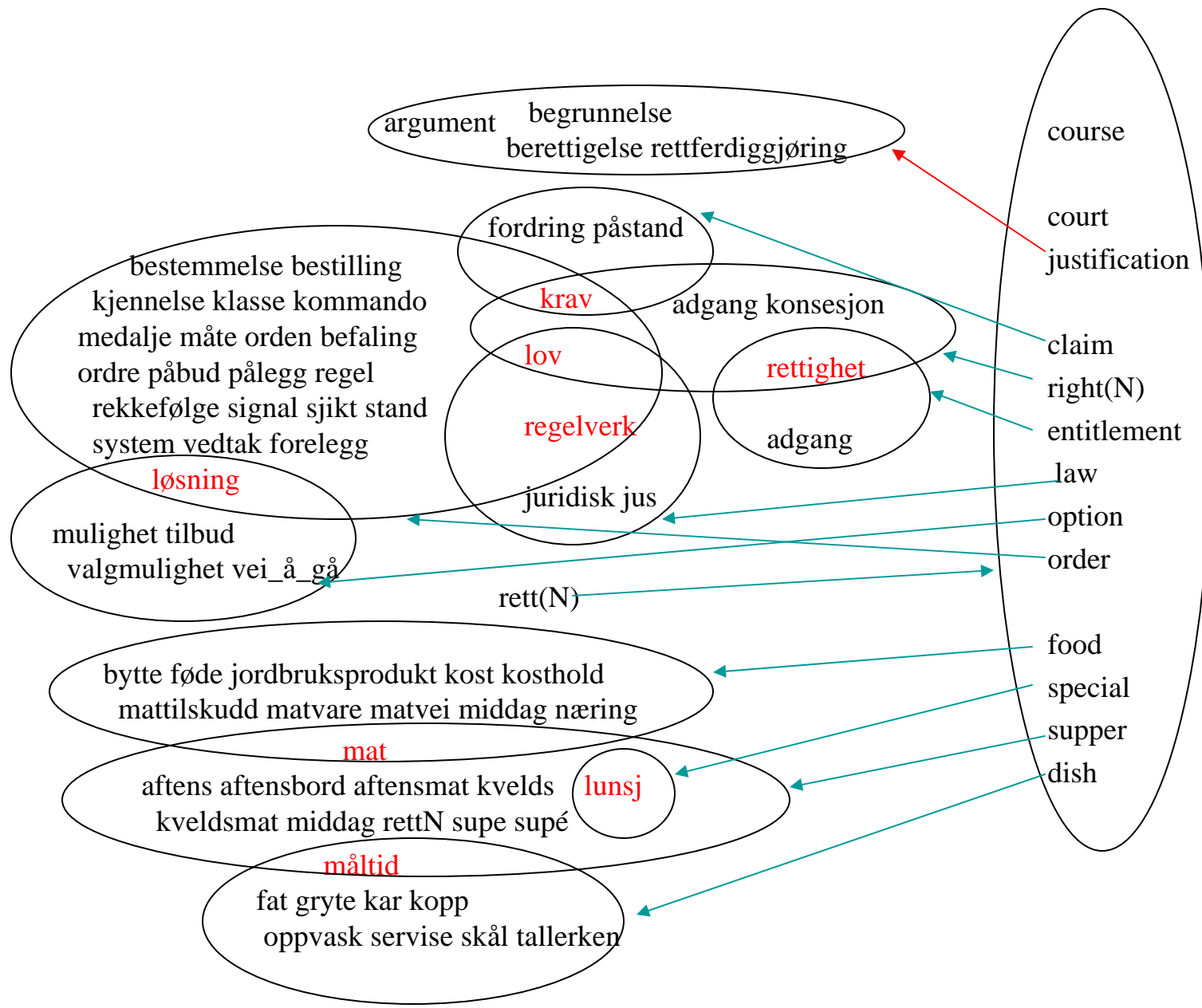


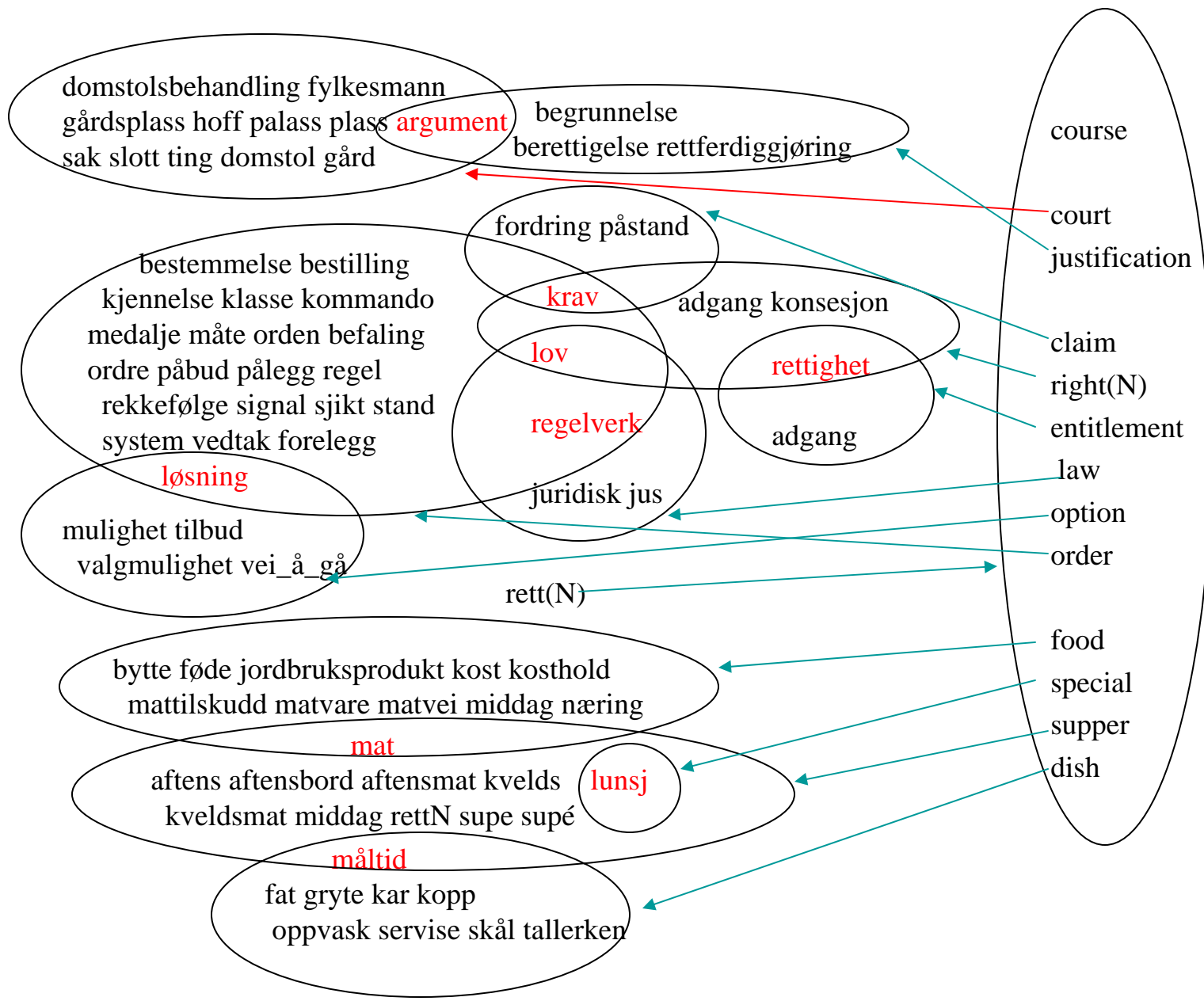


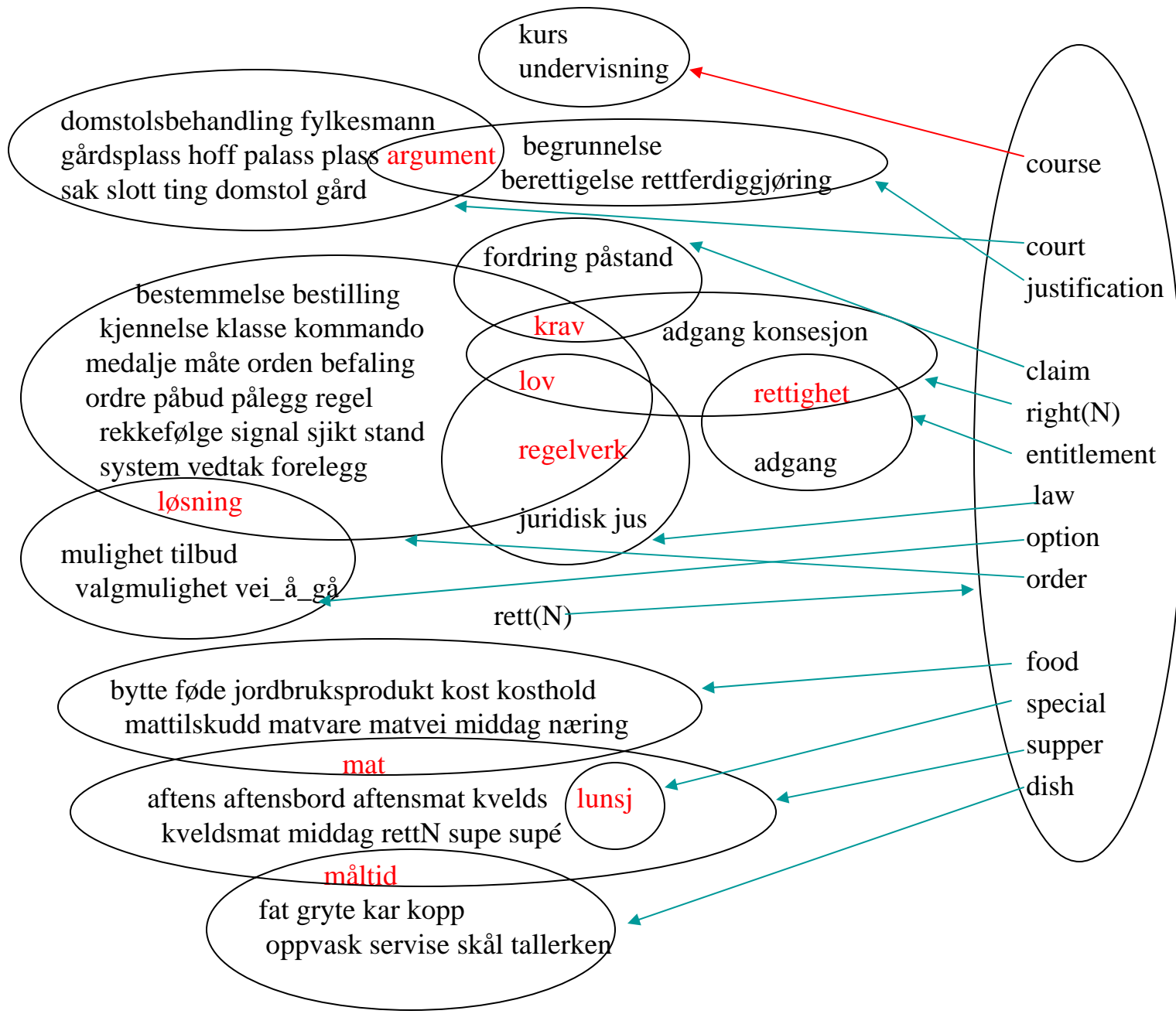


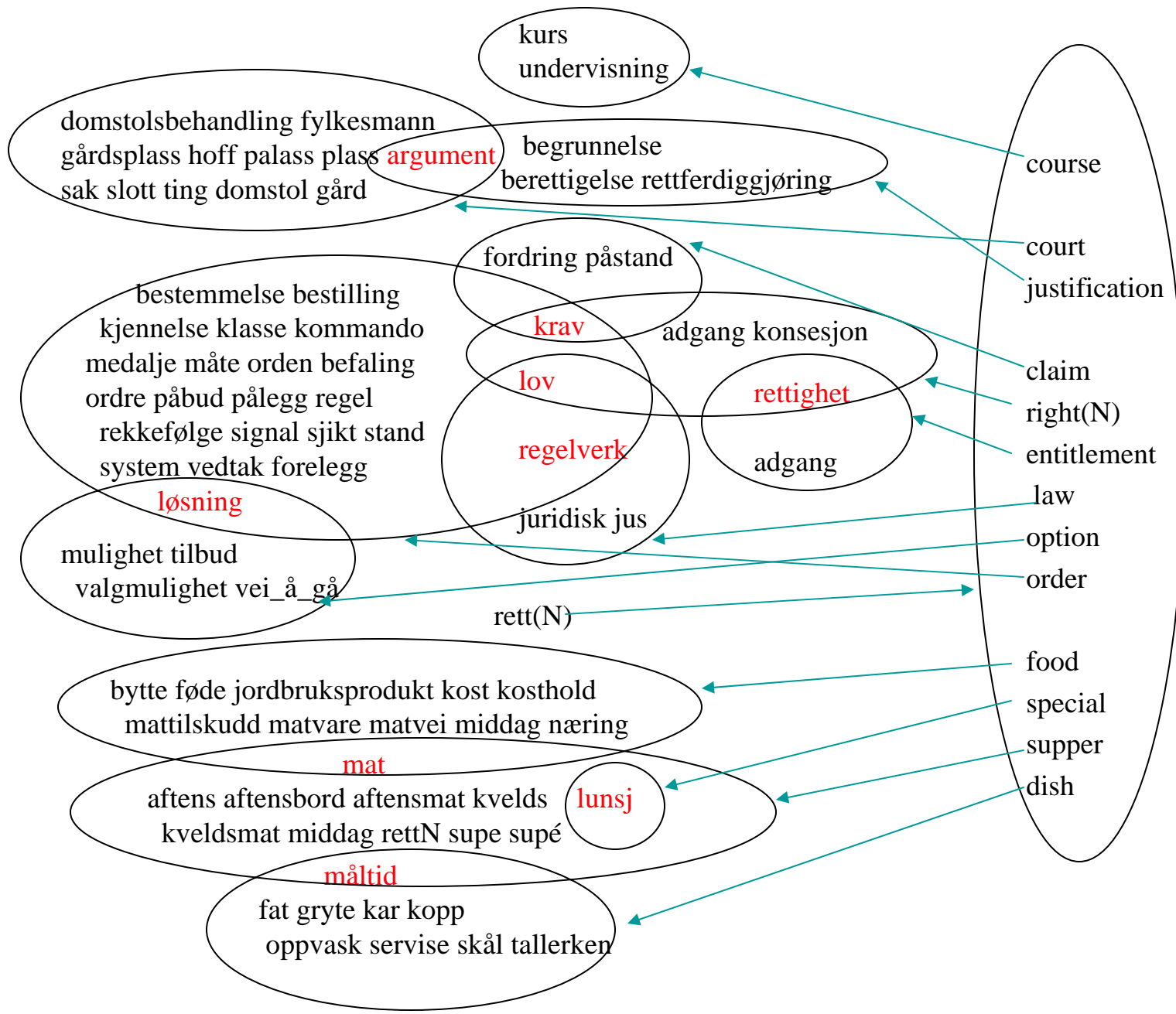


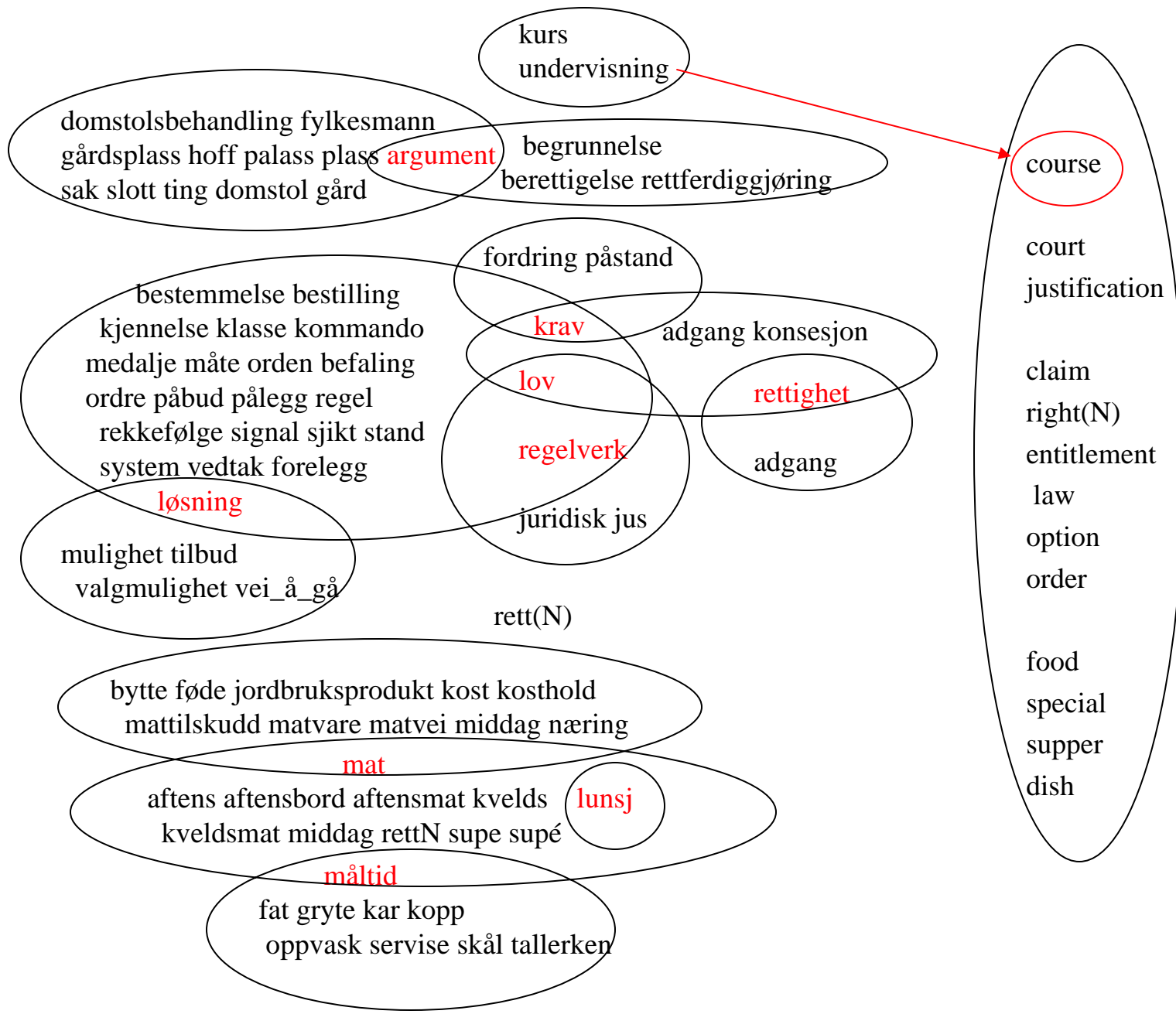


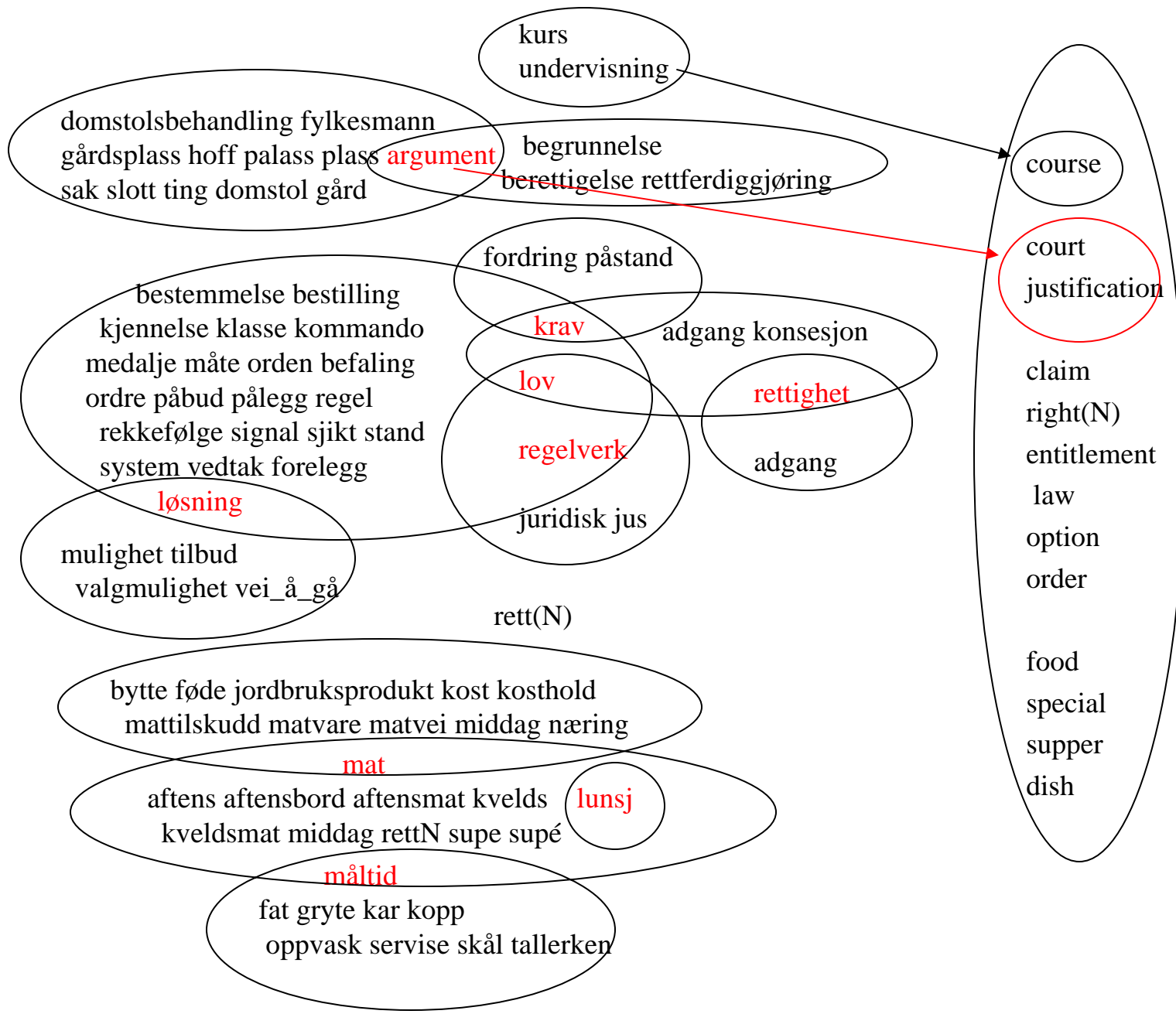


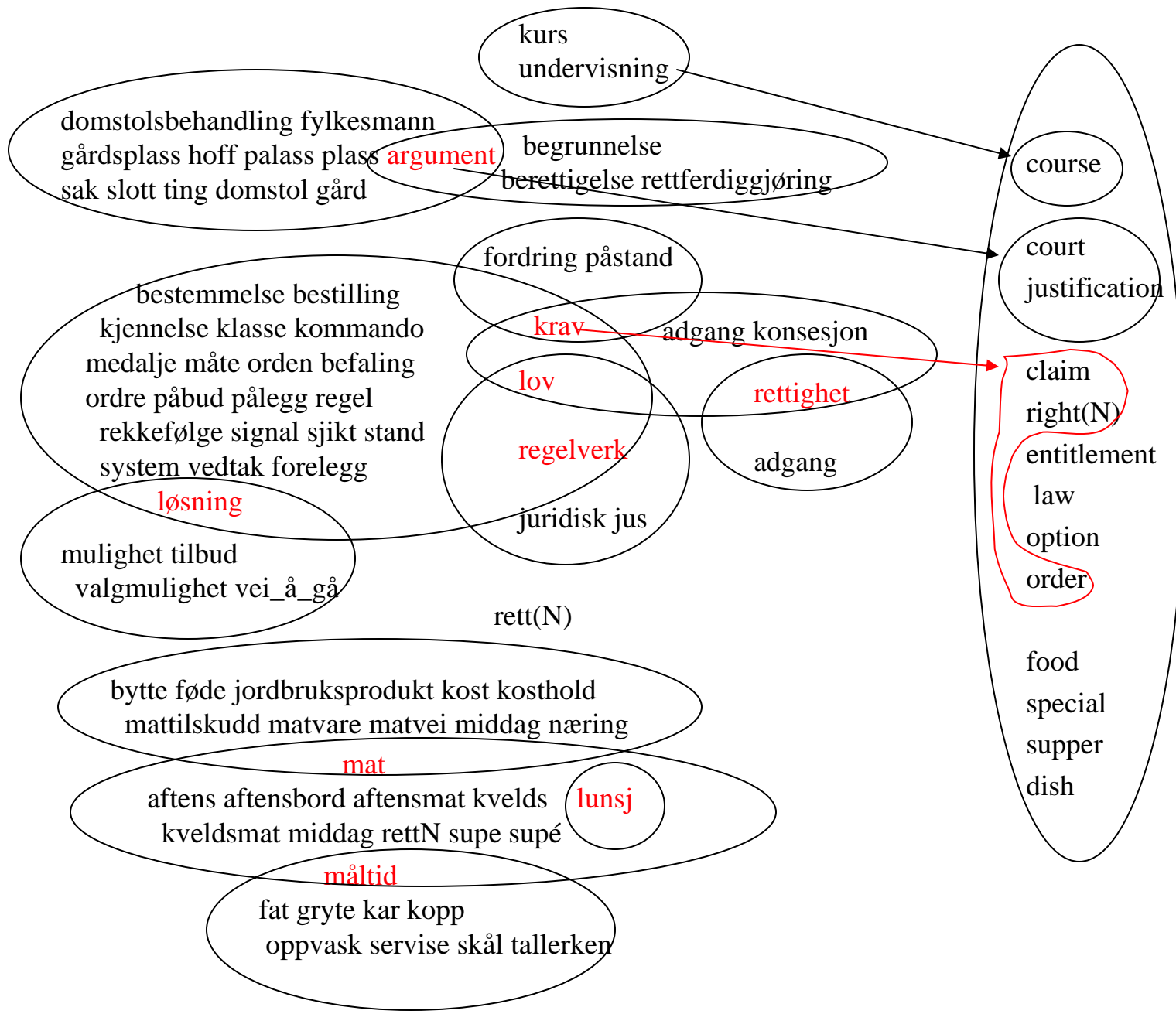


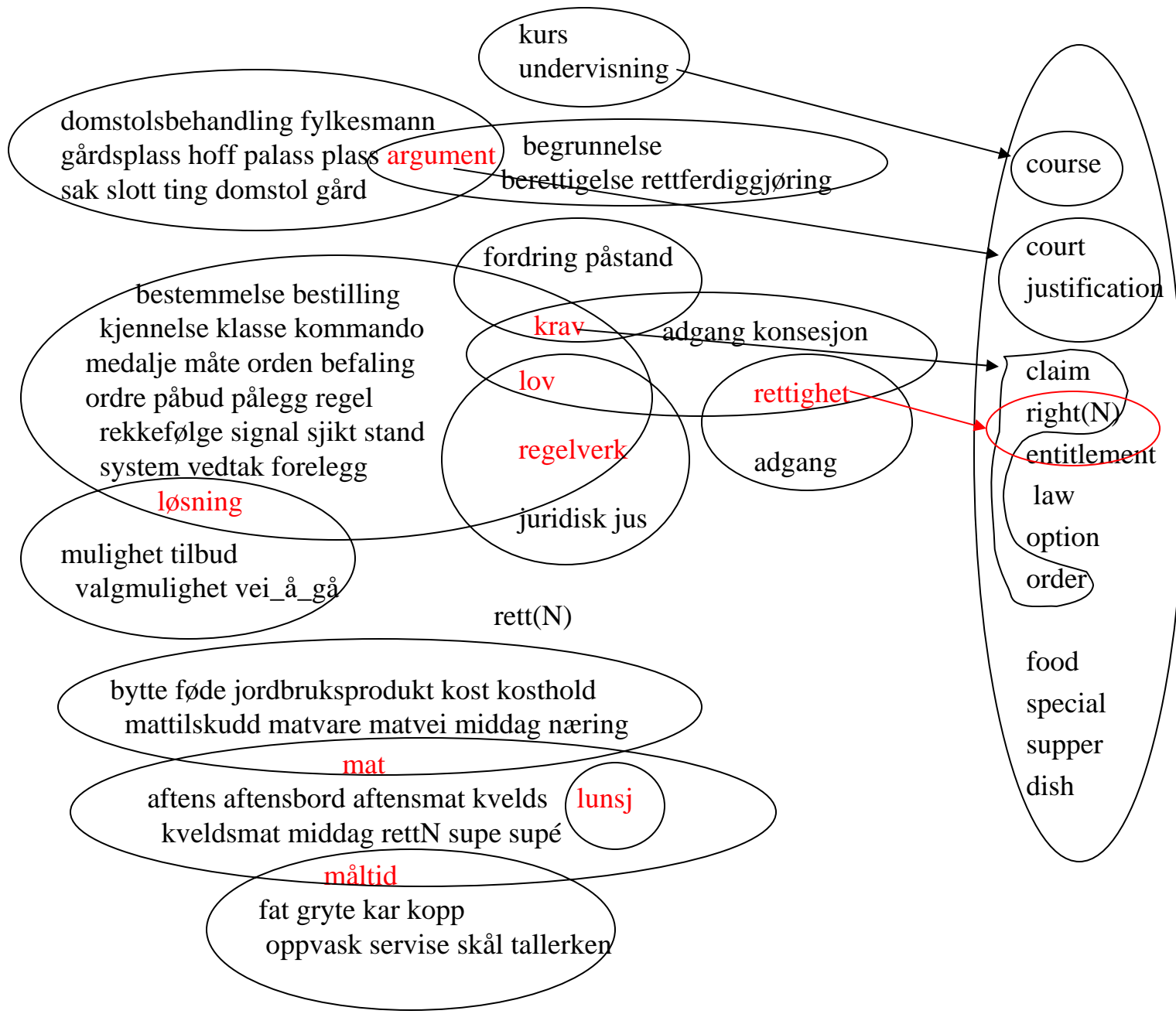


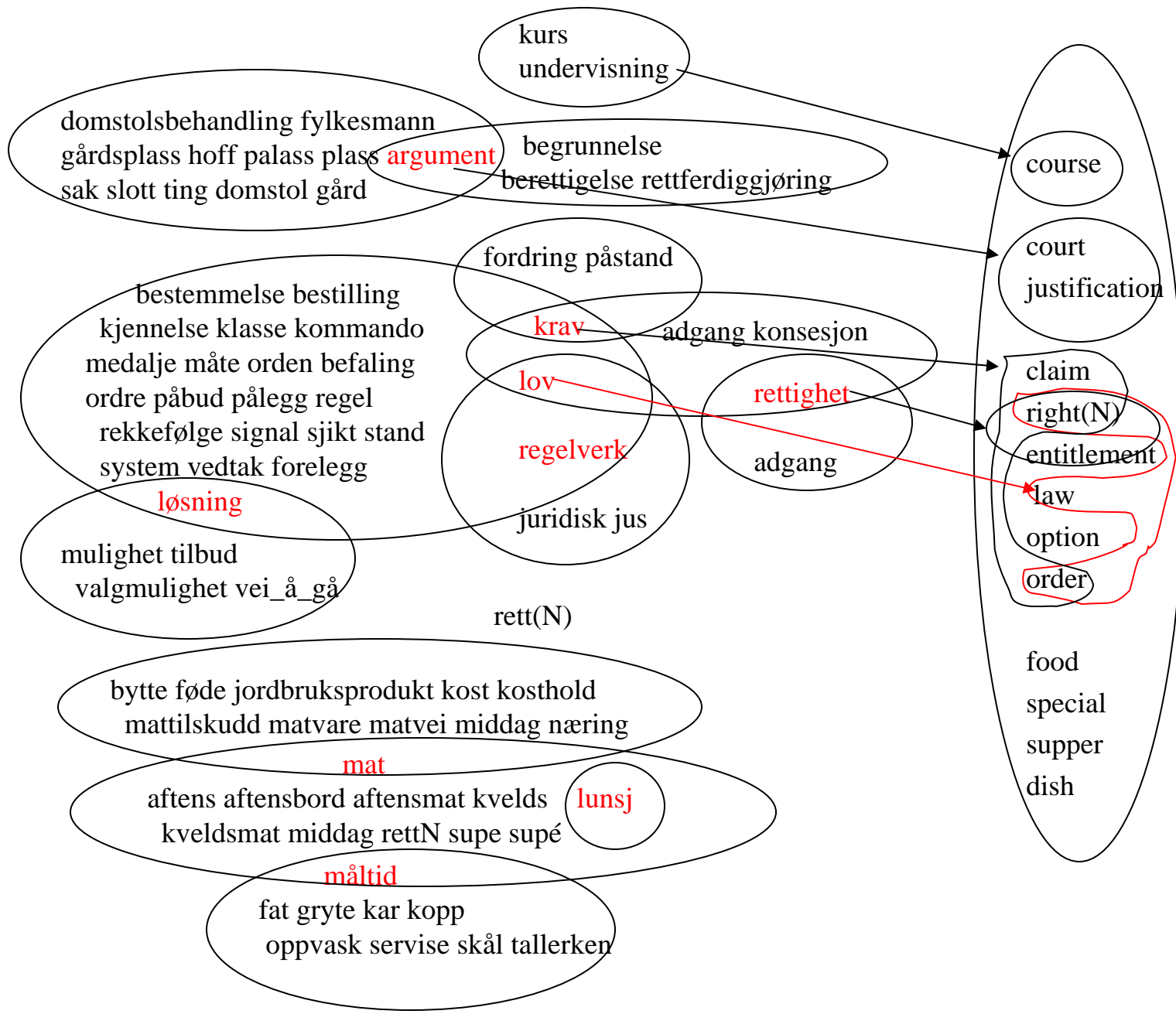


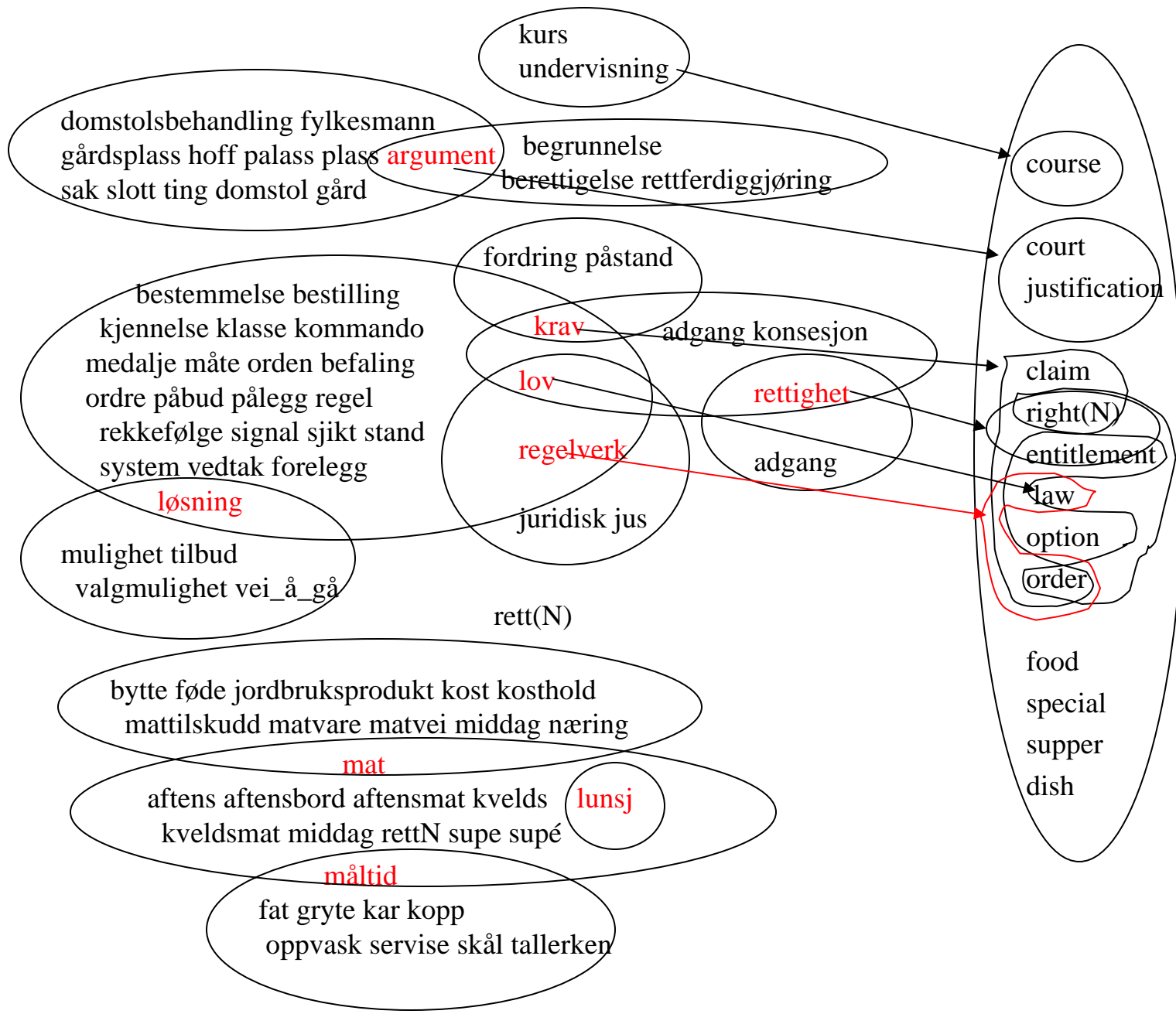


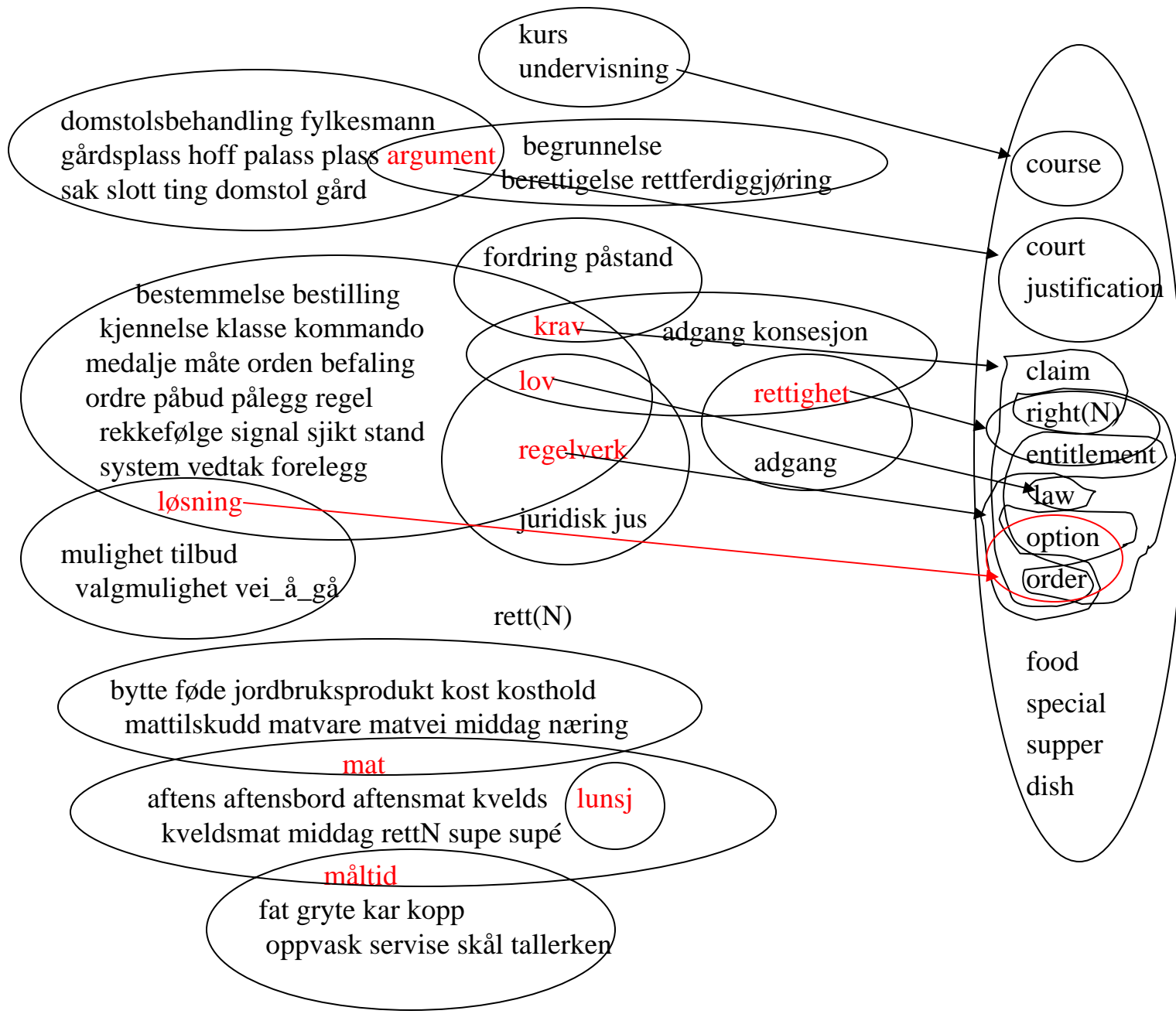


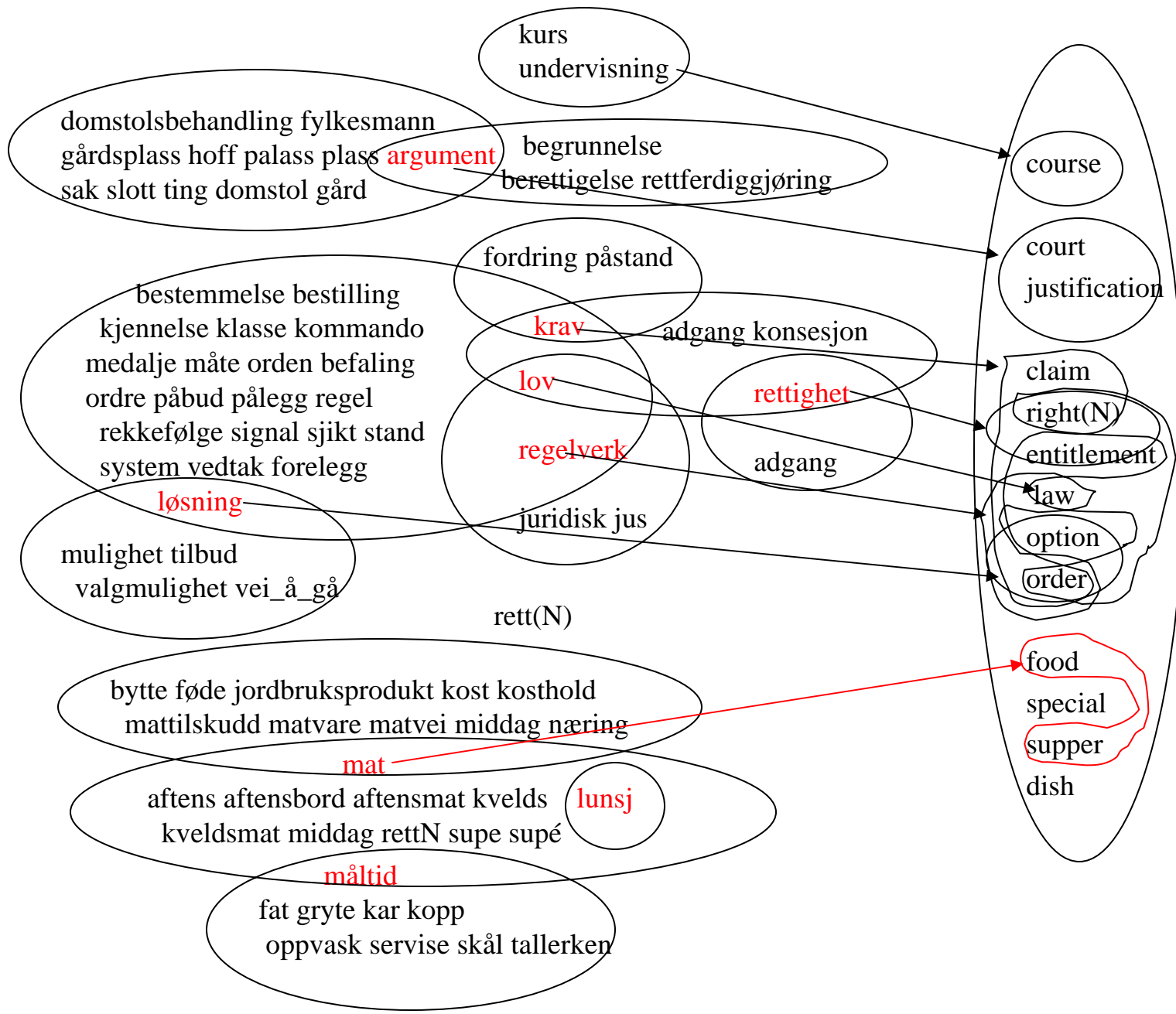


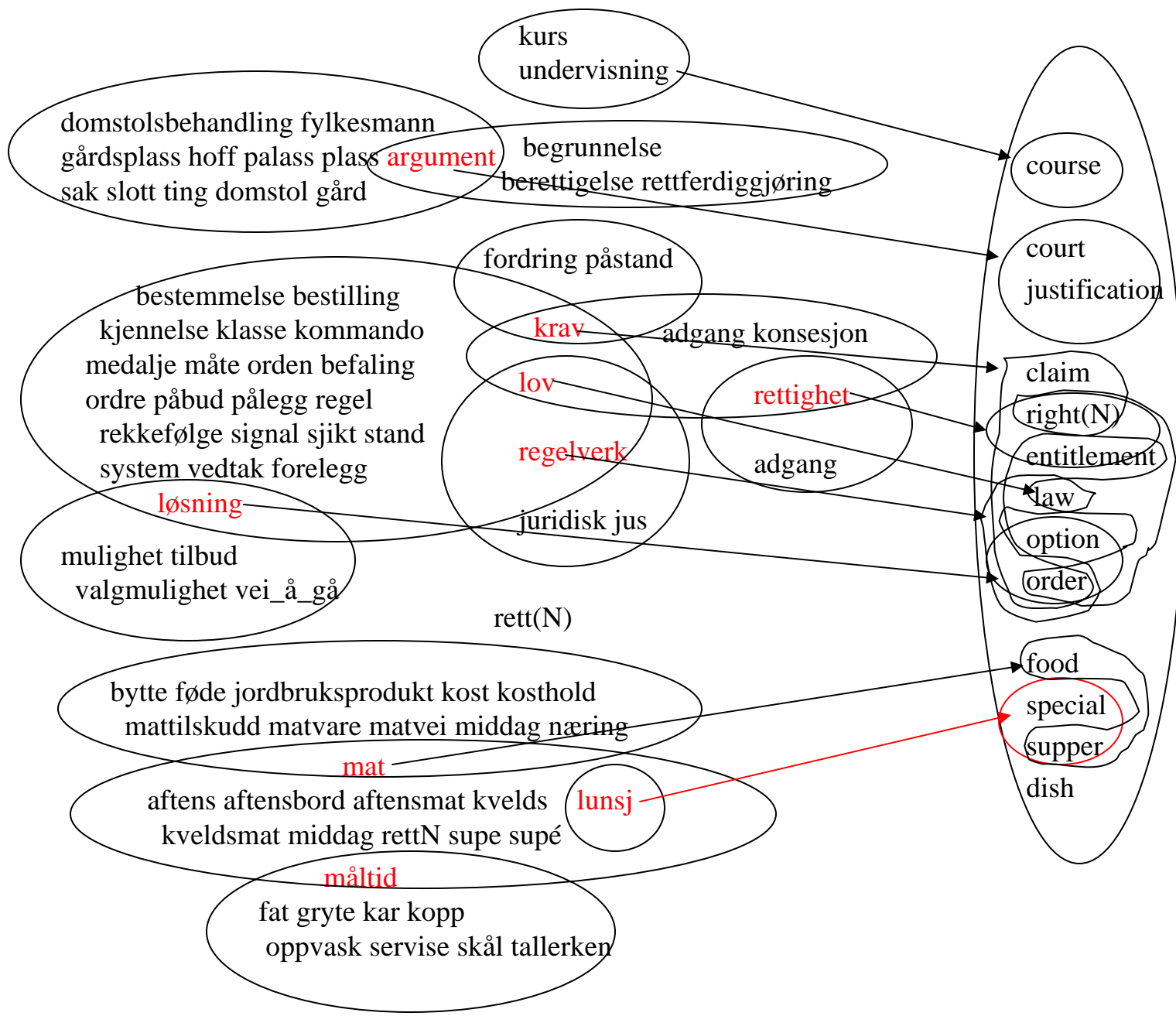


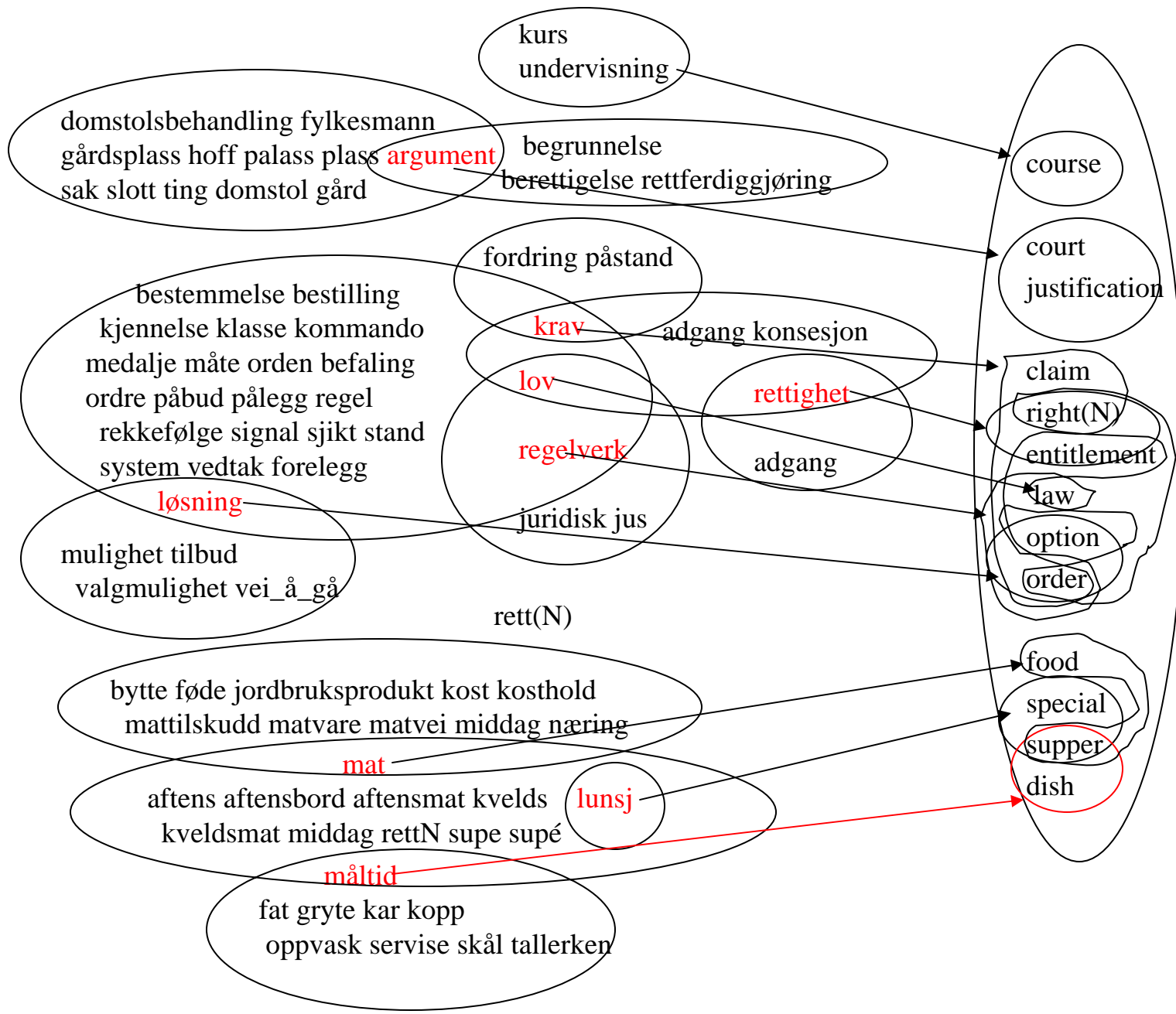




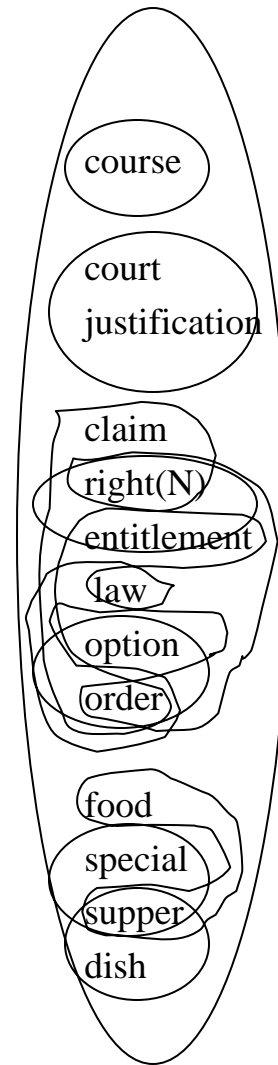


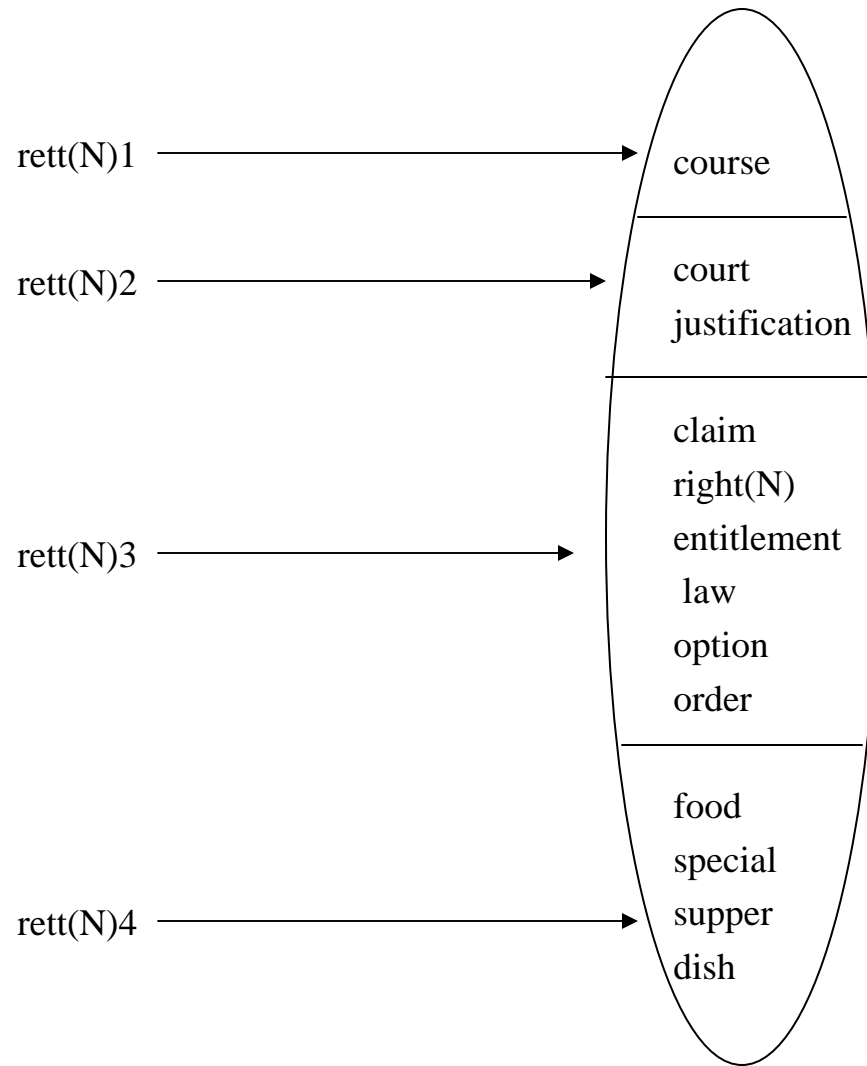






rett(N)





Semantic fields

Semantic fields

When sense individuation has been performed across the vocabulary in both languages, each sense has a *t*-image consisting of senses in the other language.

Semantic fields

When sense individuation has been performed across the vocabulary in both languages, each sense has a *t*-image consisting of senses in the other language.

On this basis senses can be collected in *semantic fields*

Semantic fields

When sense individuation has been performed across the vocabulary in both languages, each sense has a *t*-image consisting of senses in the other language.

On this basis senses can be collected in *semantic fields*

A semantic field:

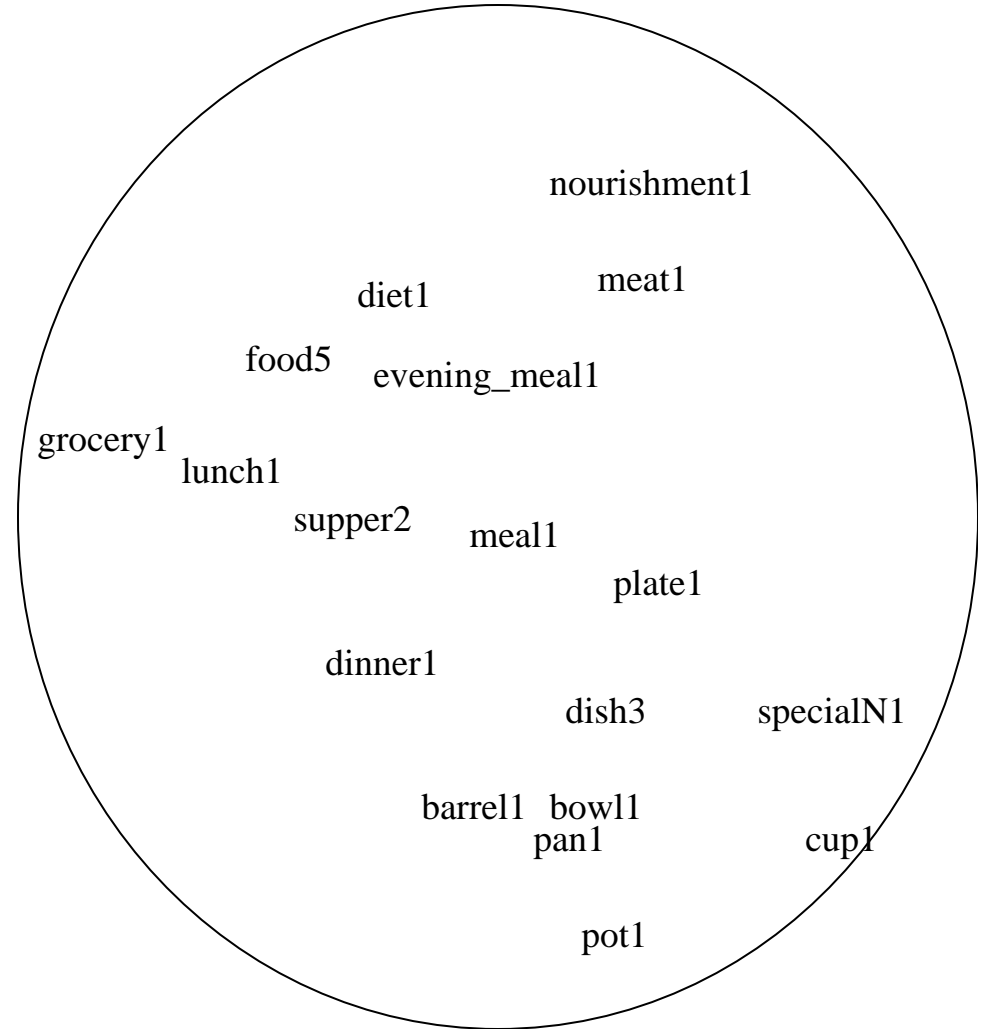
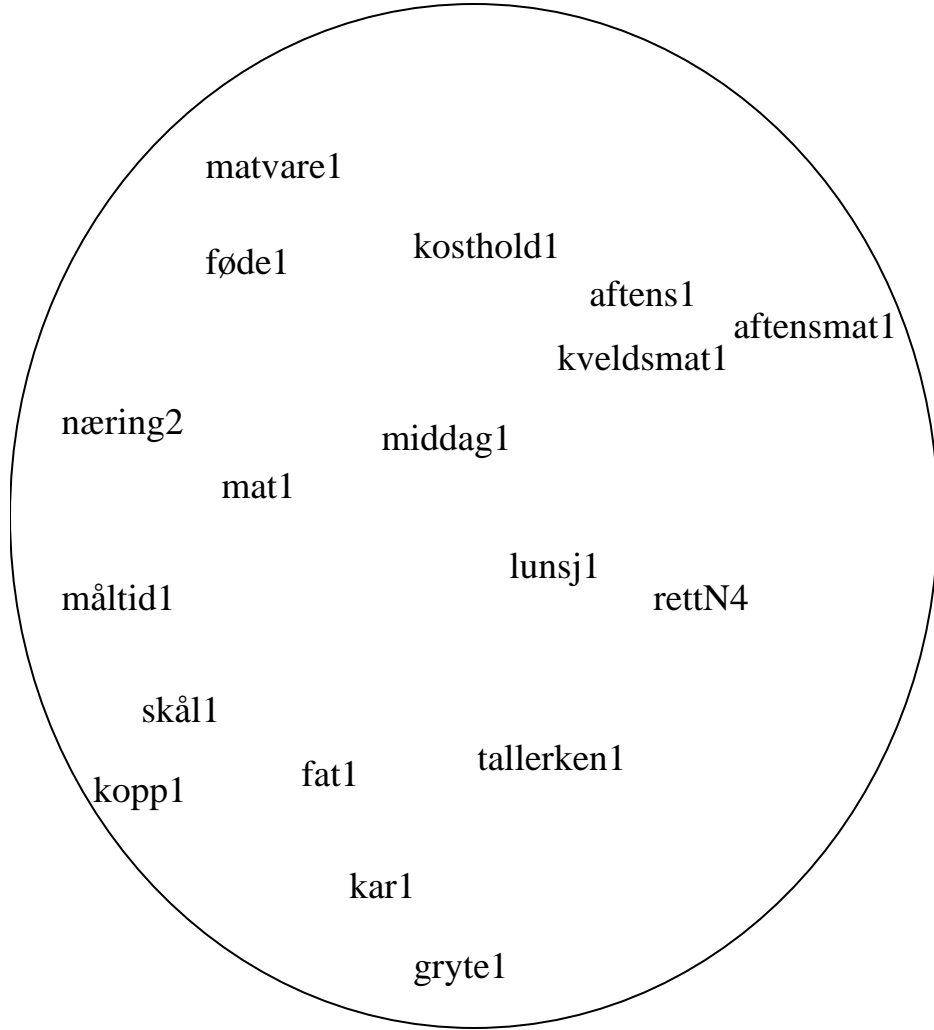
Semantic fields

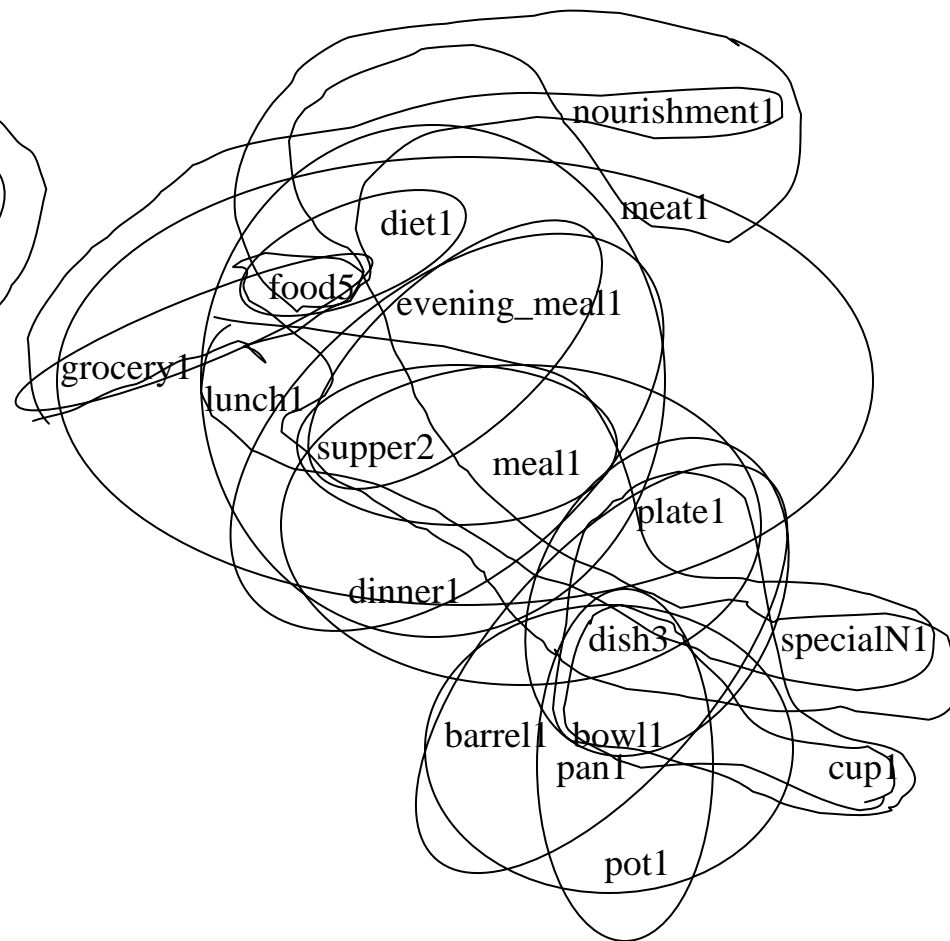
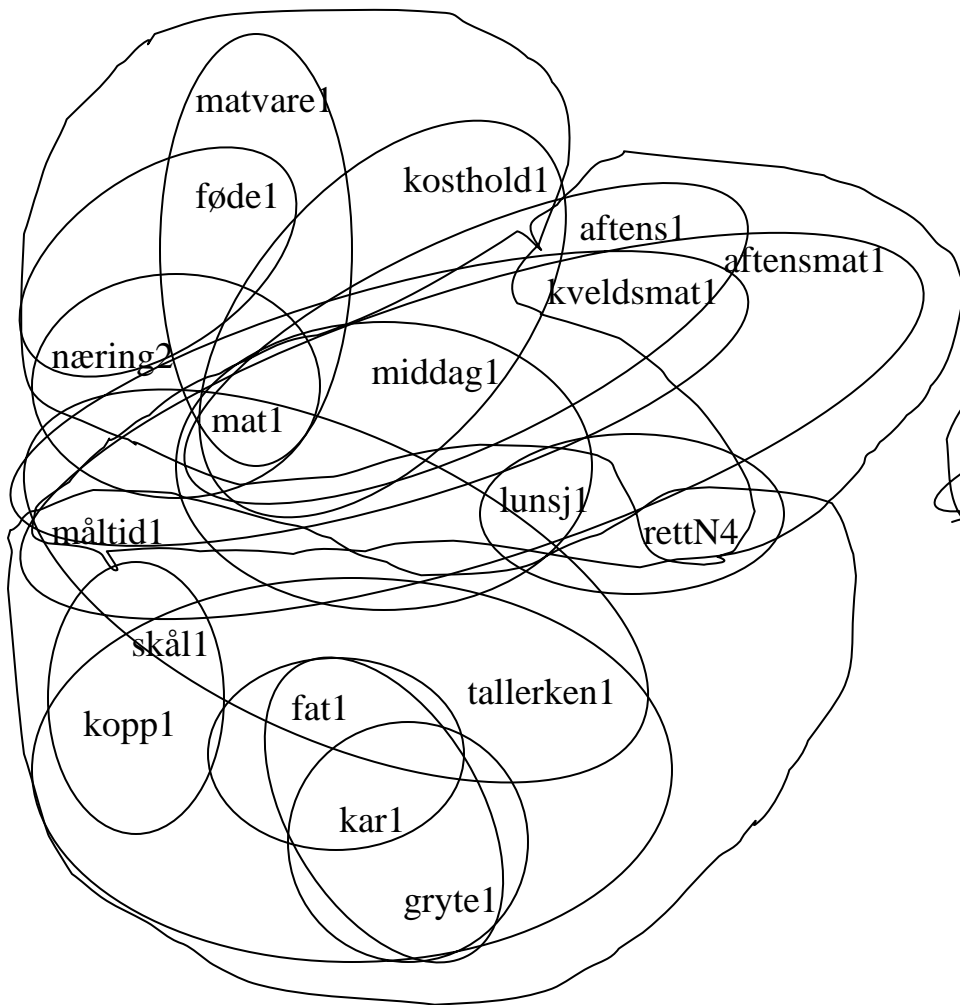
When sense individuation has been performed across the vocabulary in both languages, each sense has a *t*-image consisting of senses in the other language.

On this basis senses can be collected in *semantic fields*

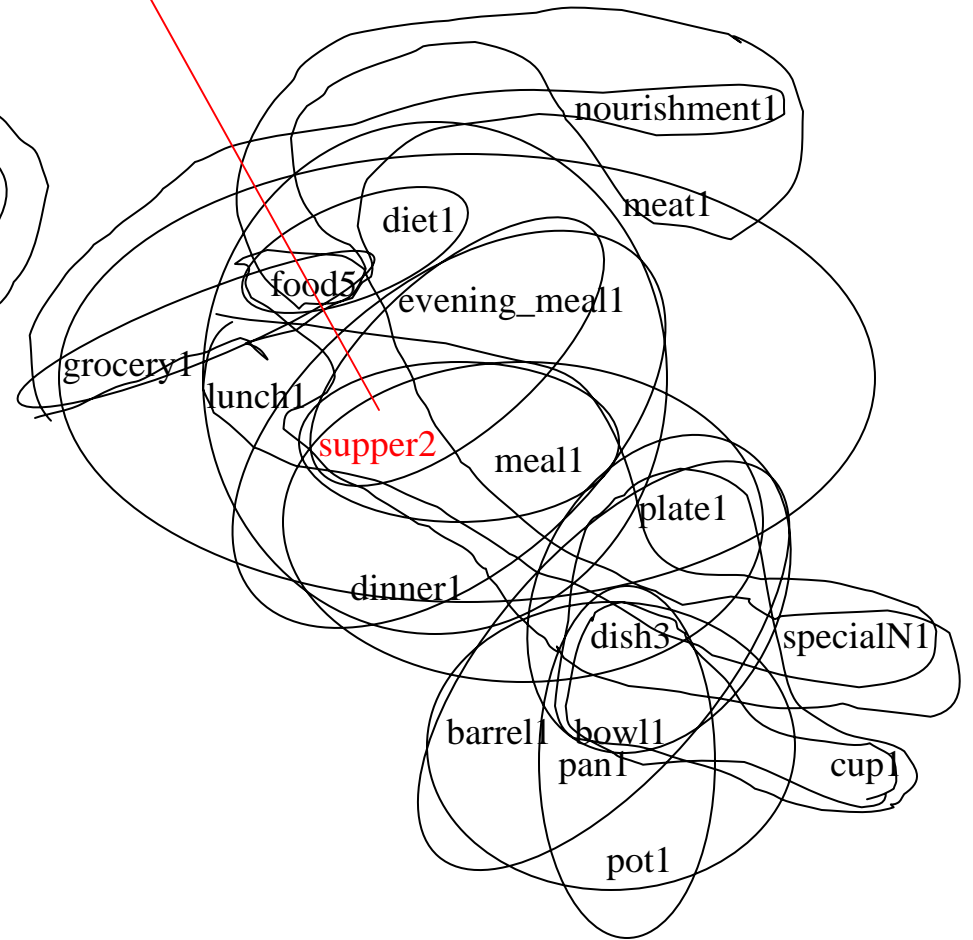
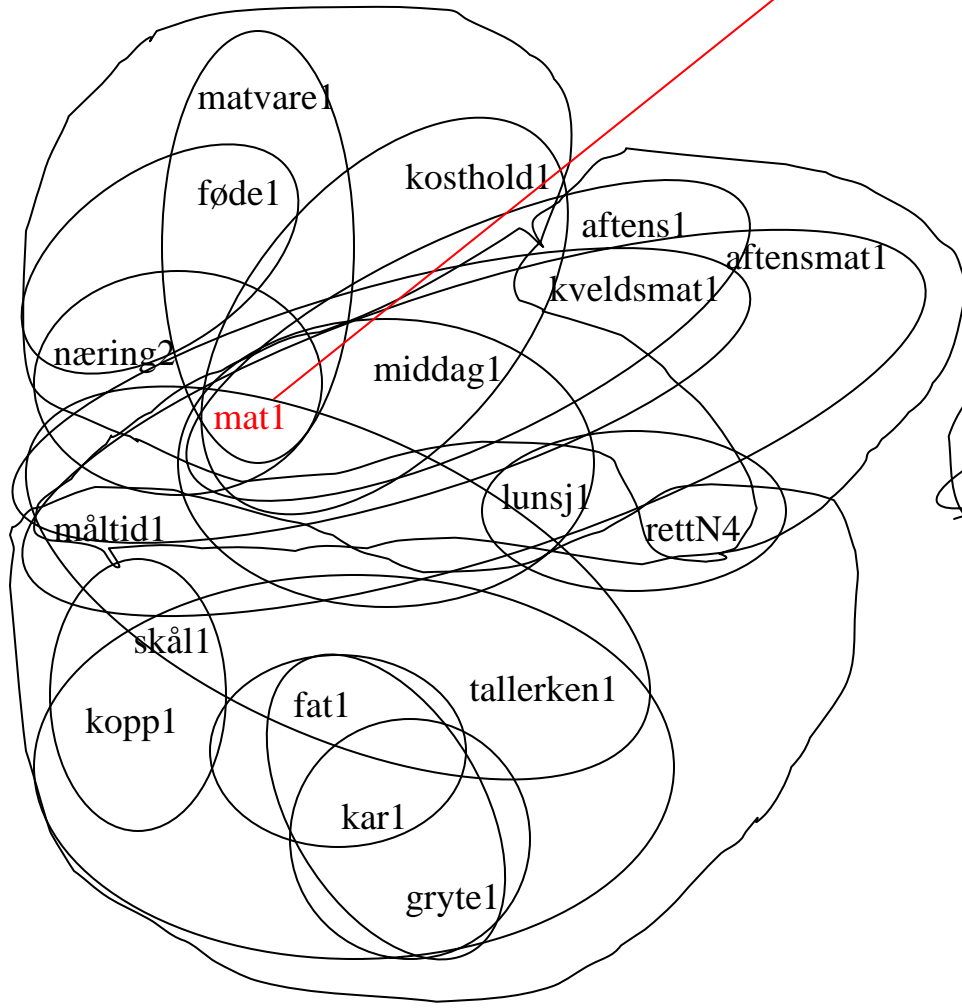
A semantic field:

A set of word senses with directly or indirectly intersecting *t*-images

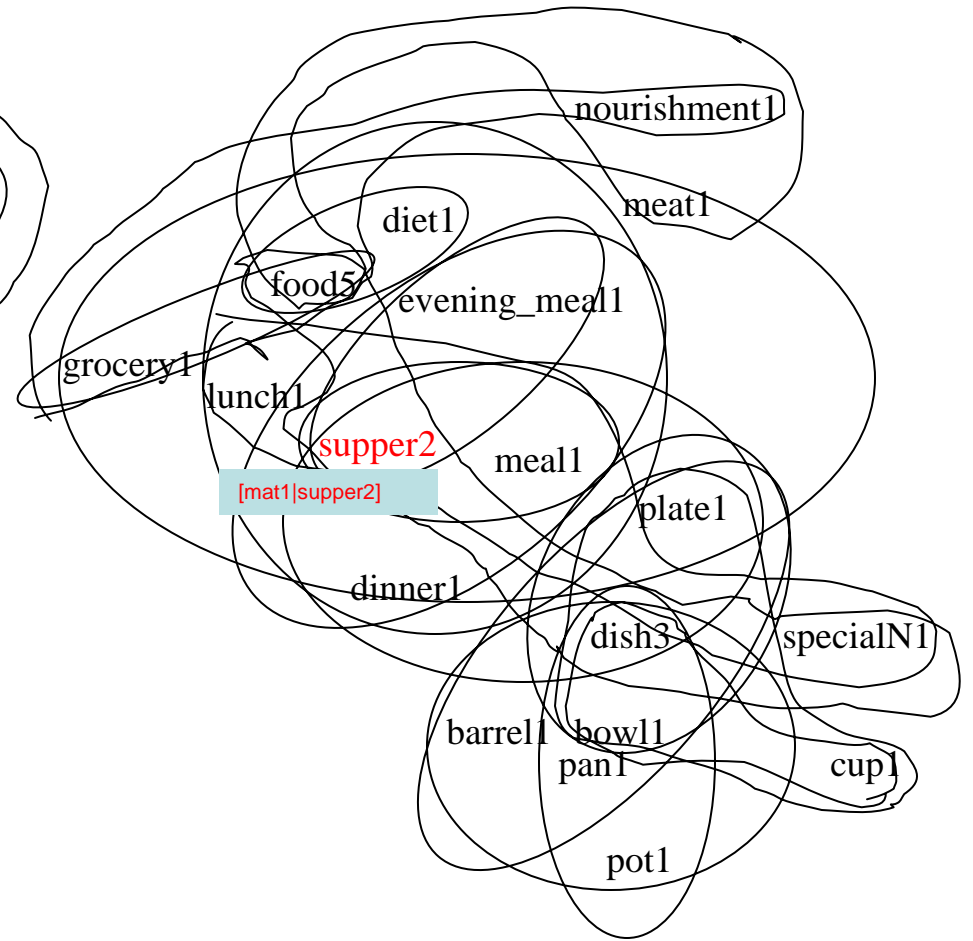
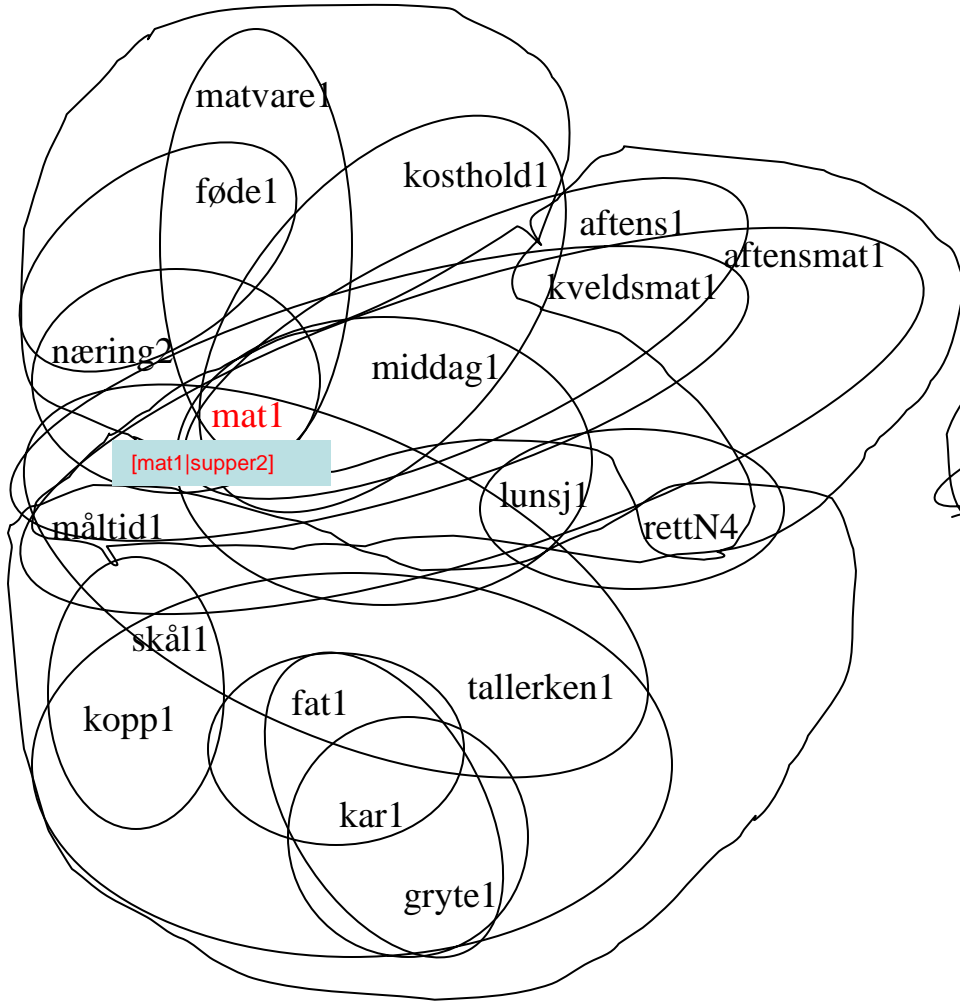




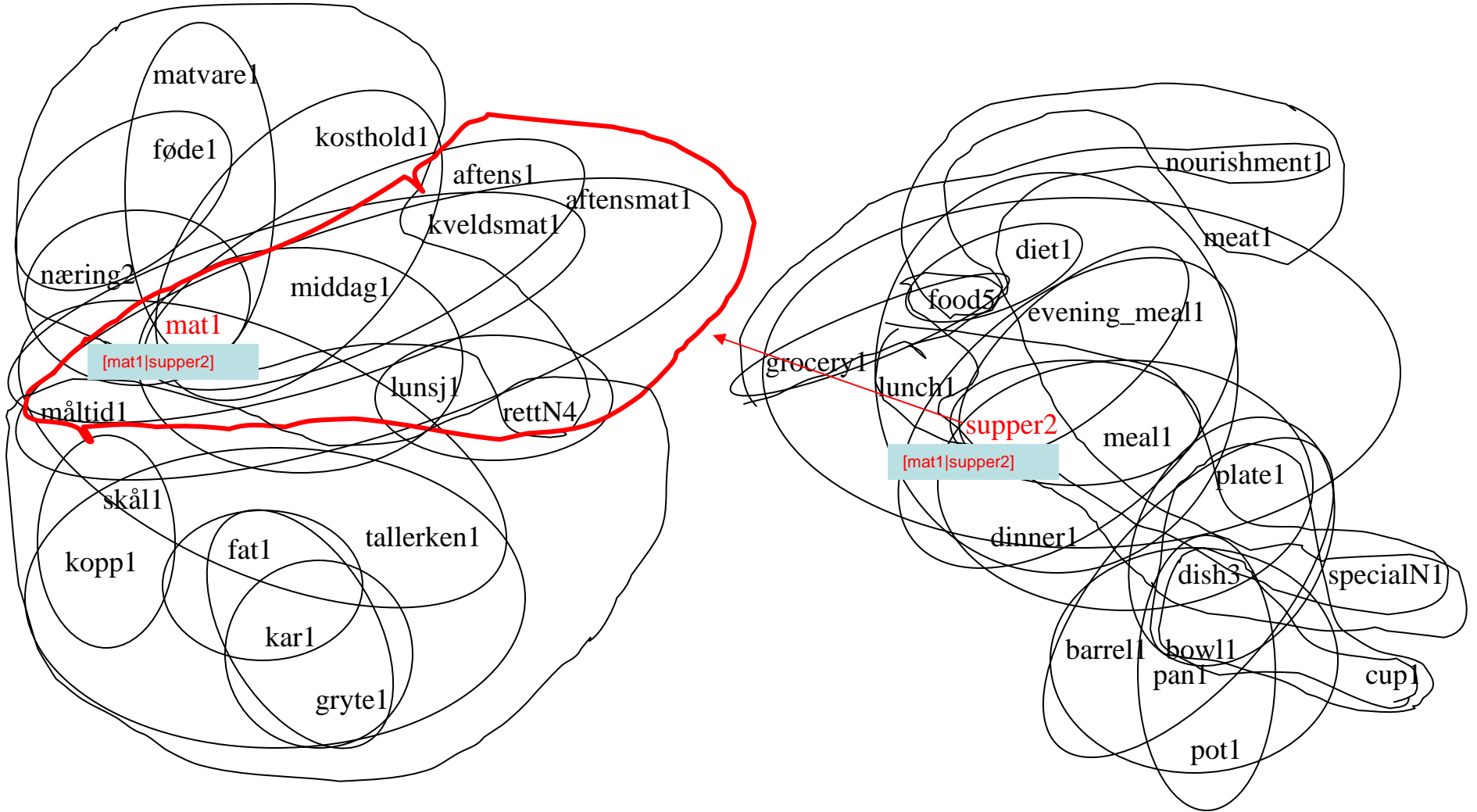
[mat1|supper2]



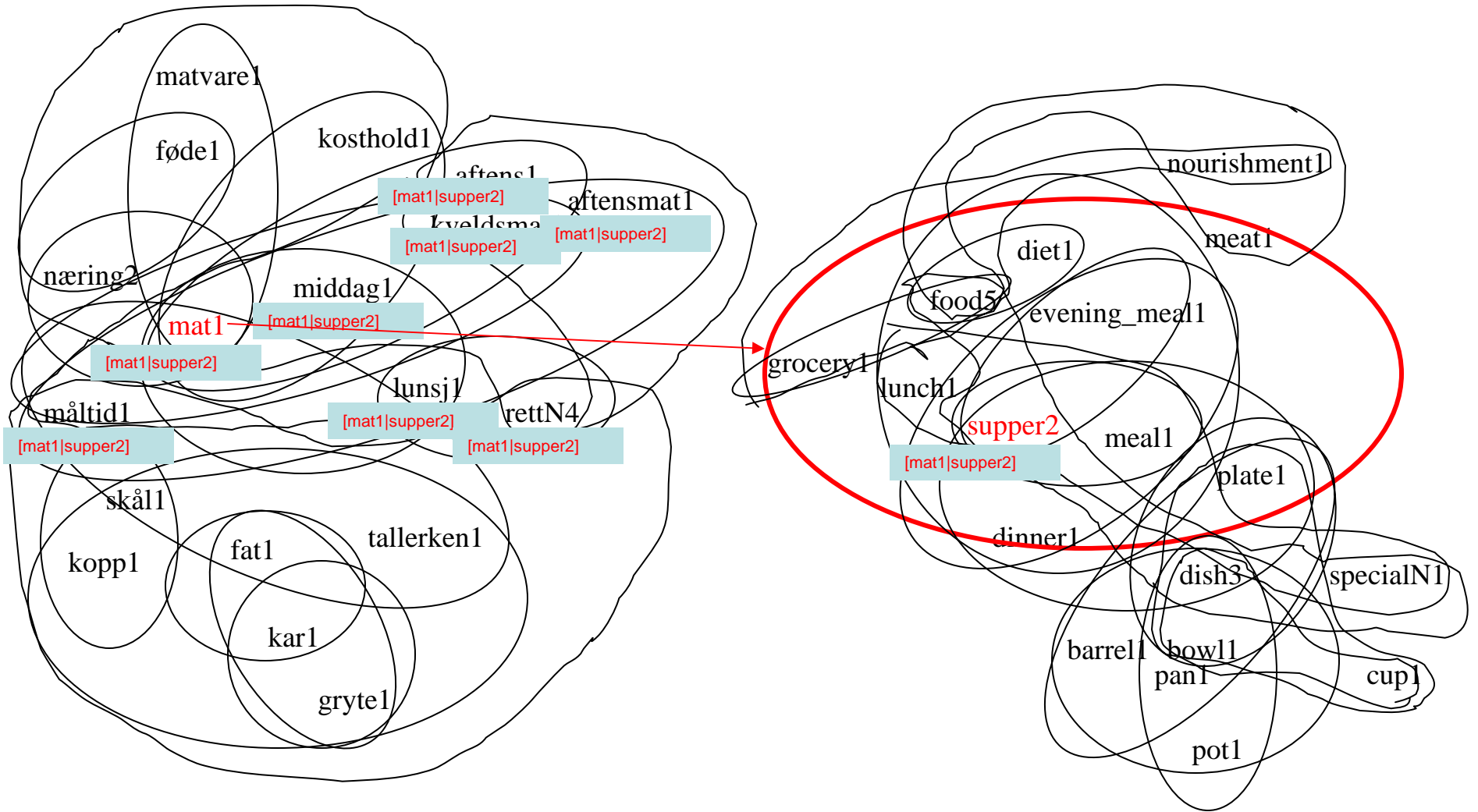
[mat1|supper2]



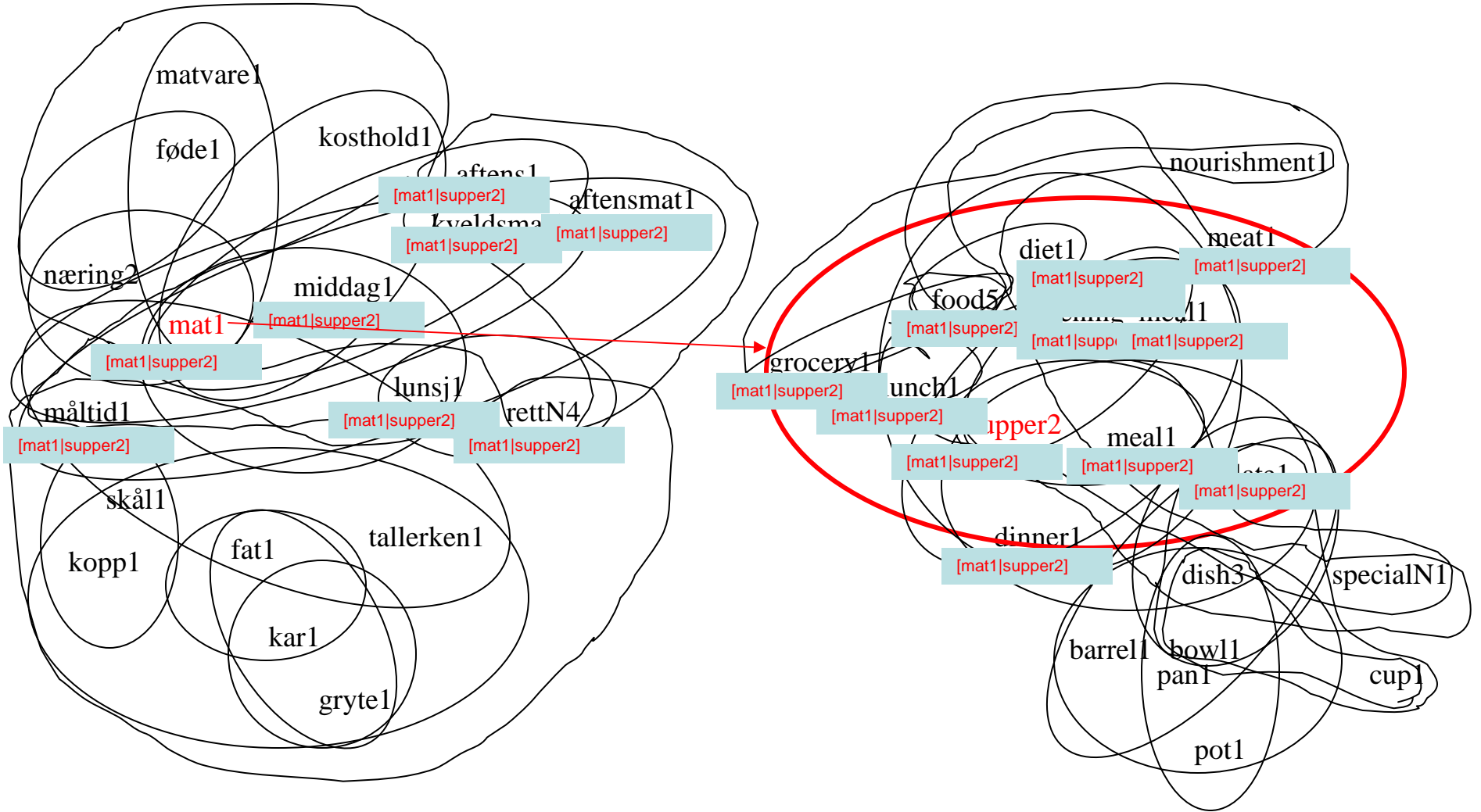
[mat1|supper2]



[mat1|supper2]

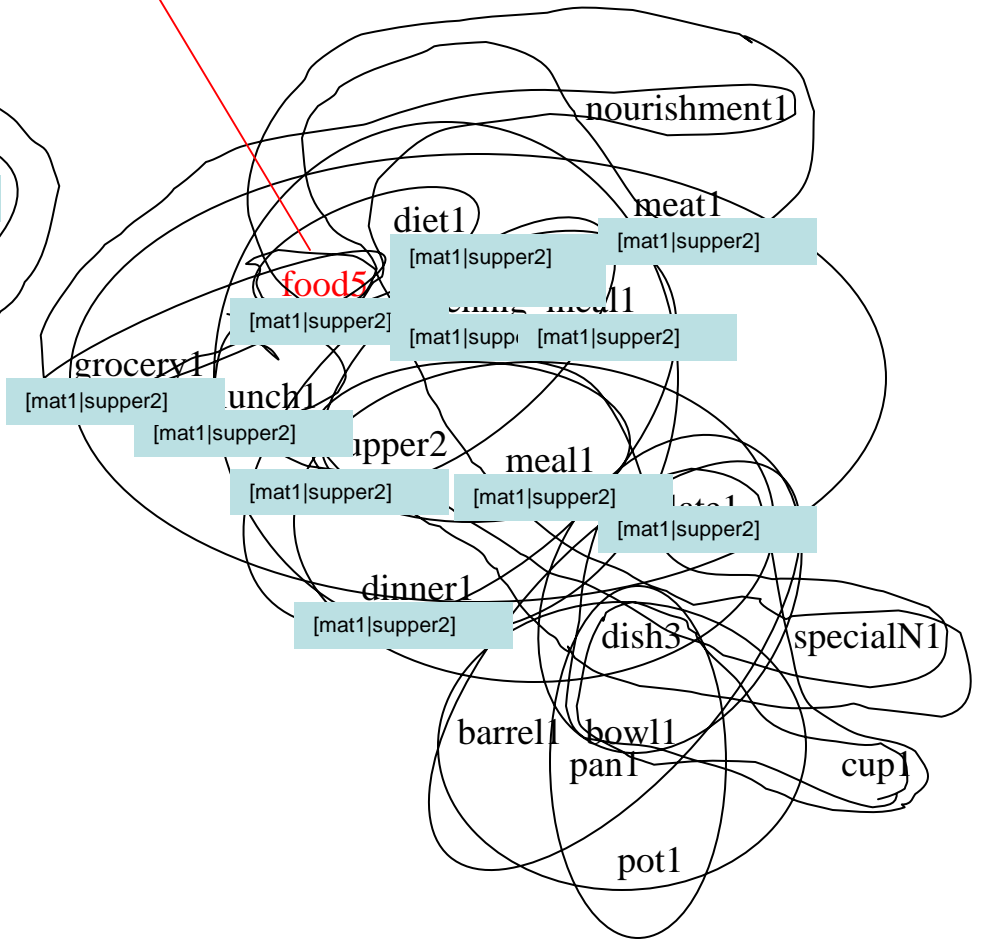
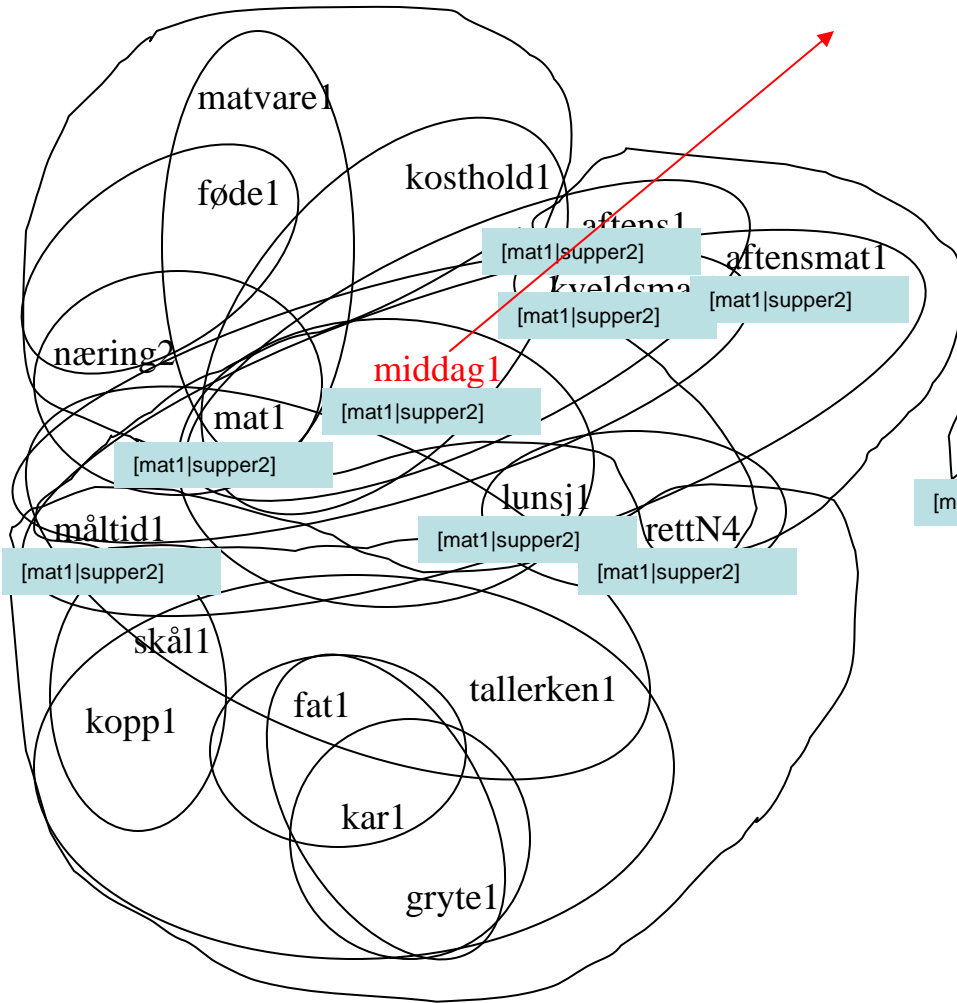


[mat1|supper2]



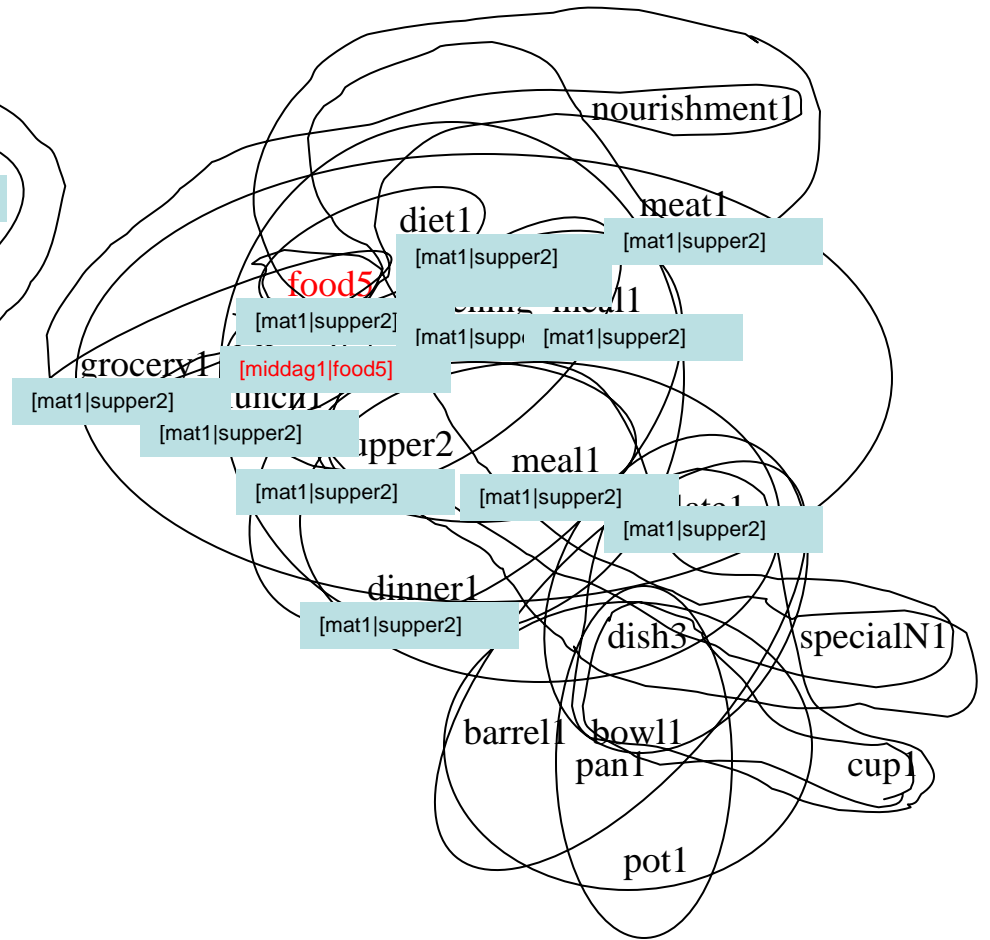
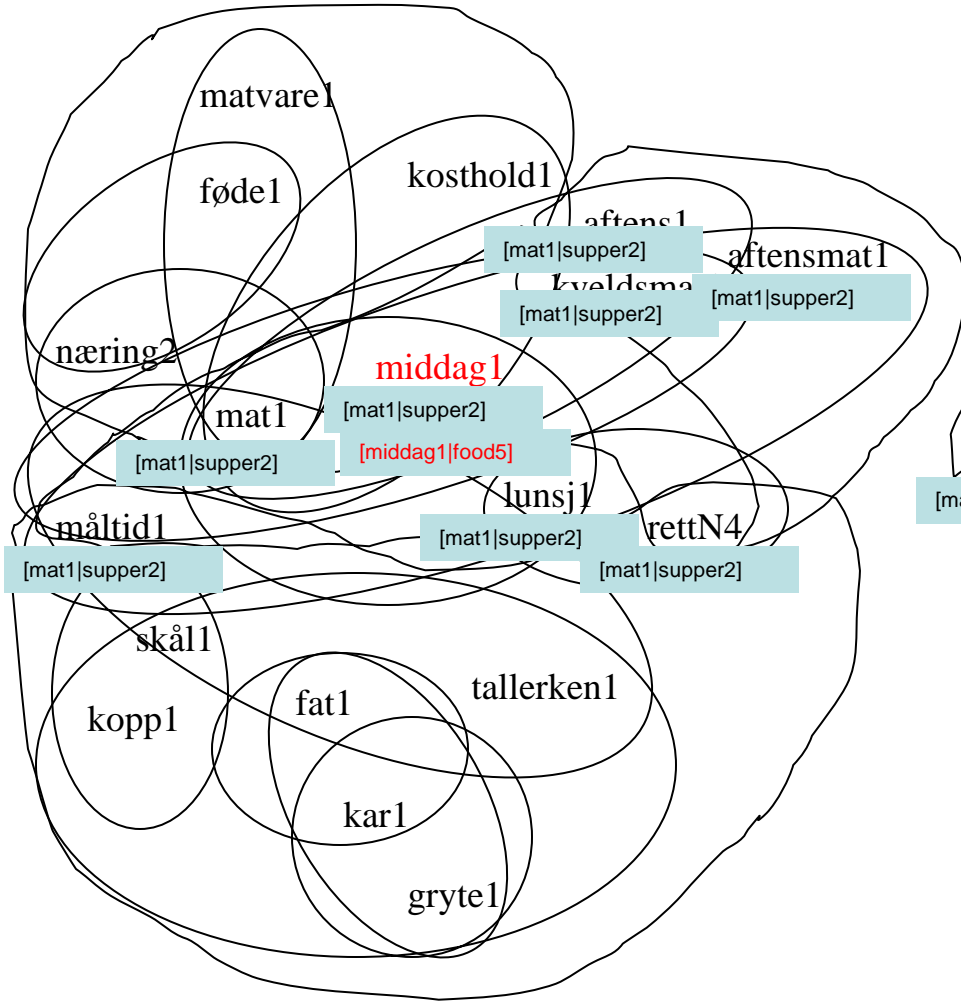
[mat1|supper2]

[middag1|food5]



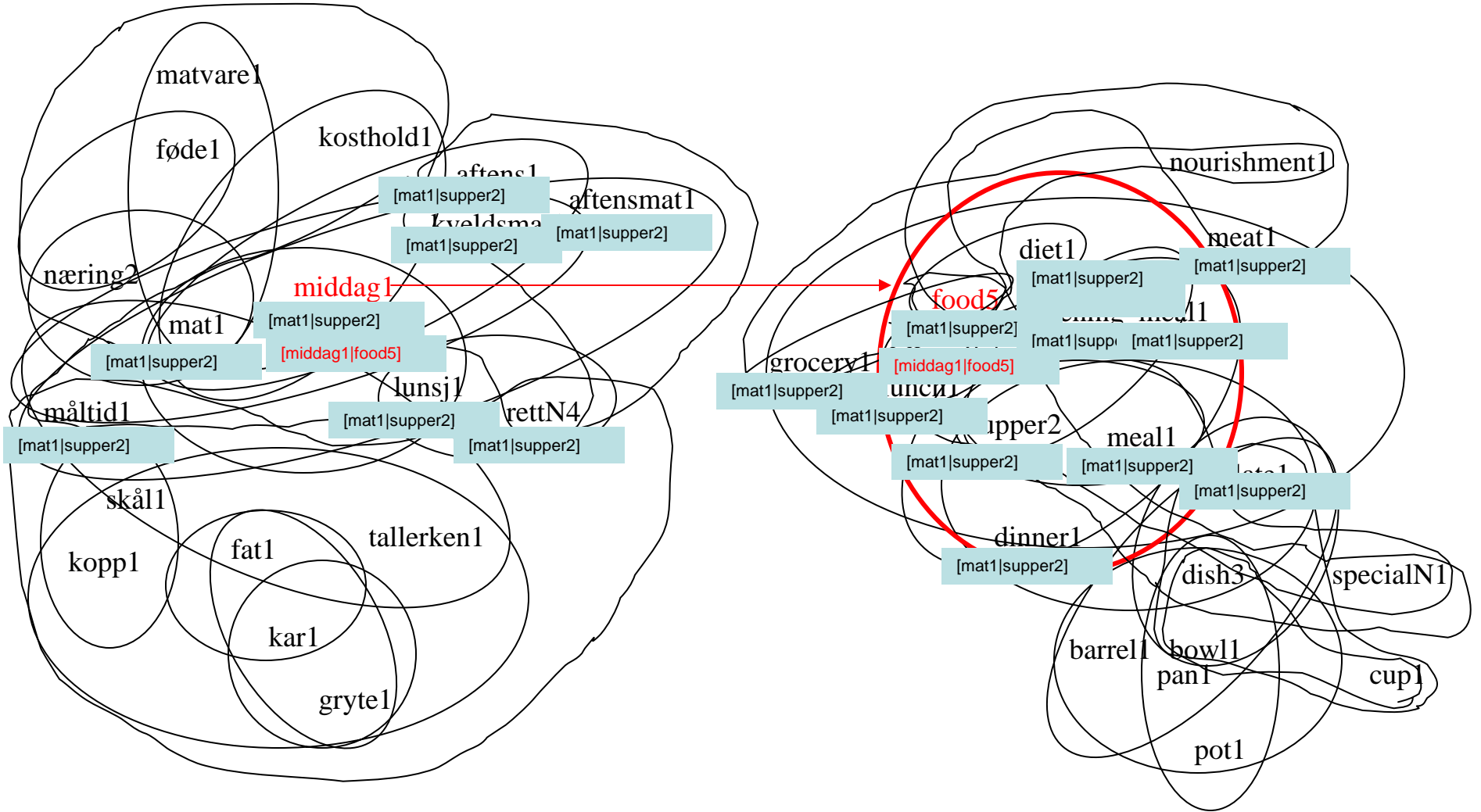
[mat1|supper2]

[middag1|food5]



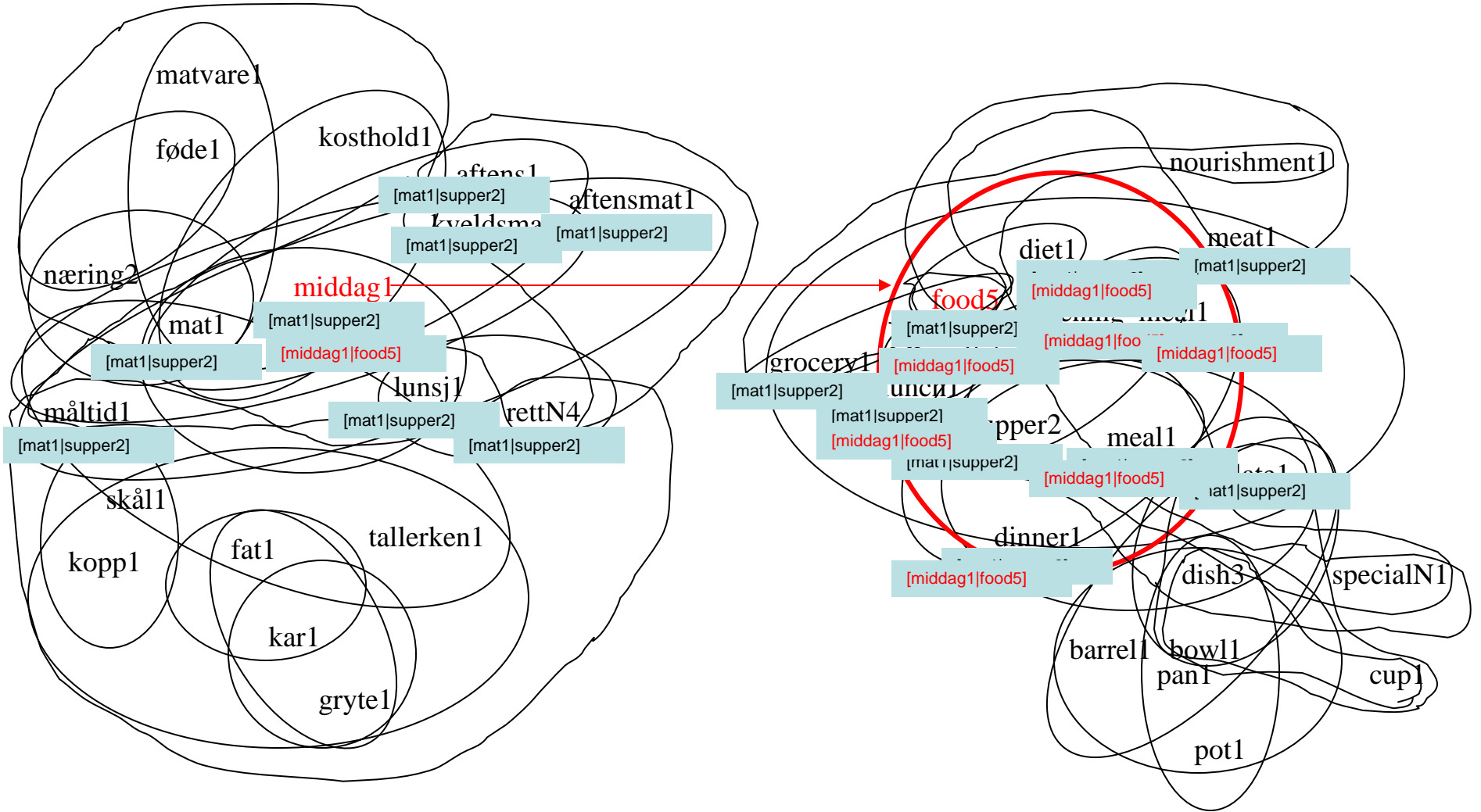
[mat1|supper2]

[middag1|food5]



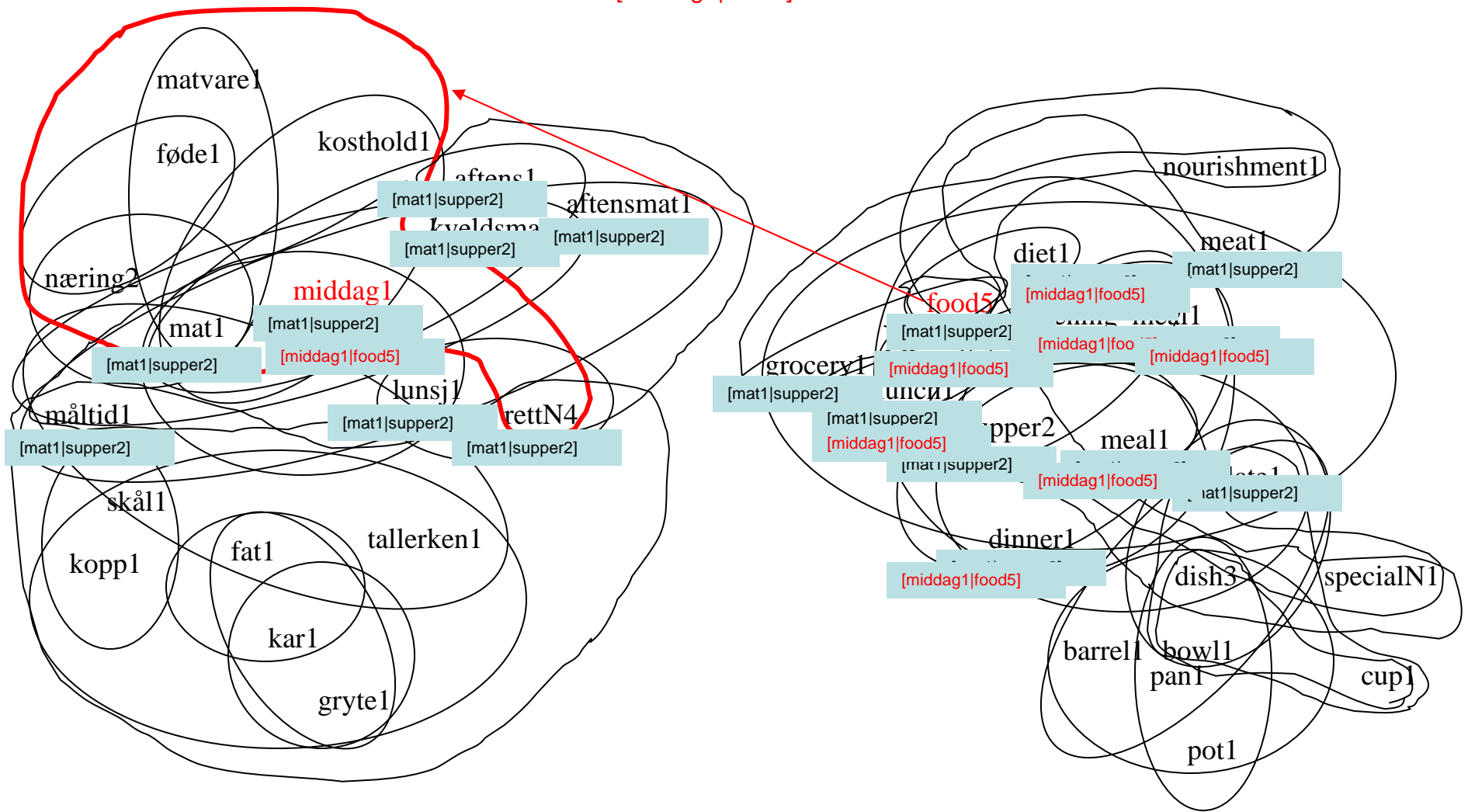
[mat1|supper2]

[middag1|food5]



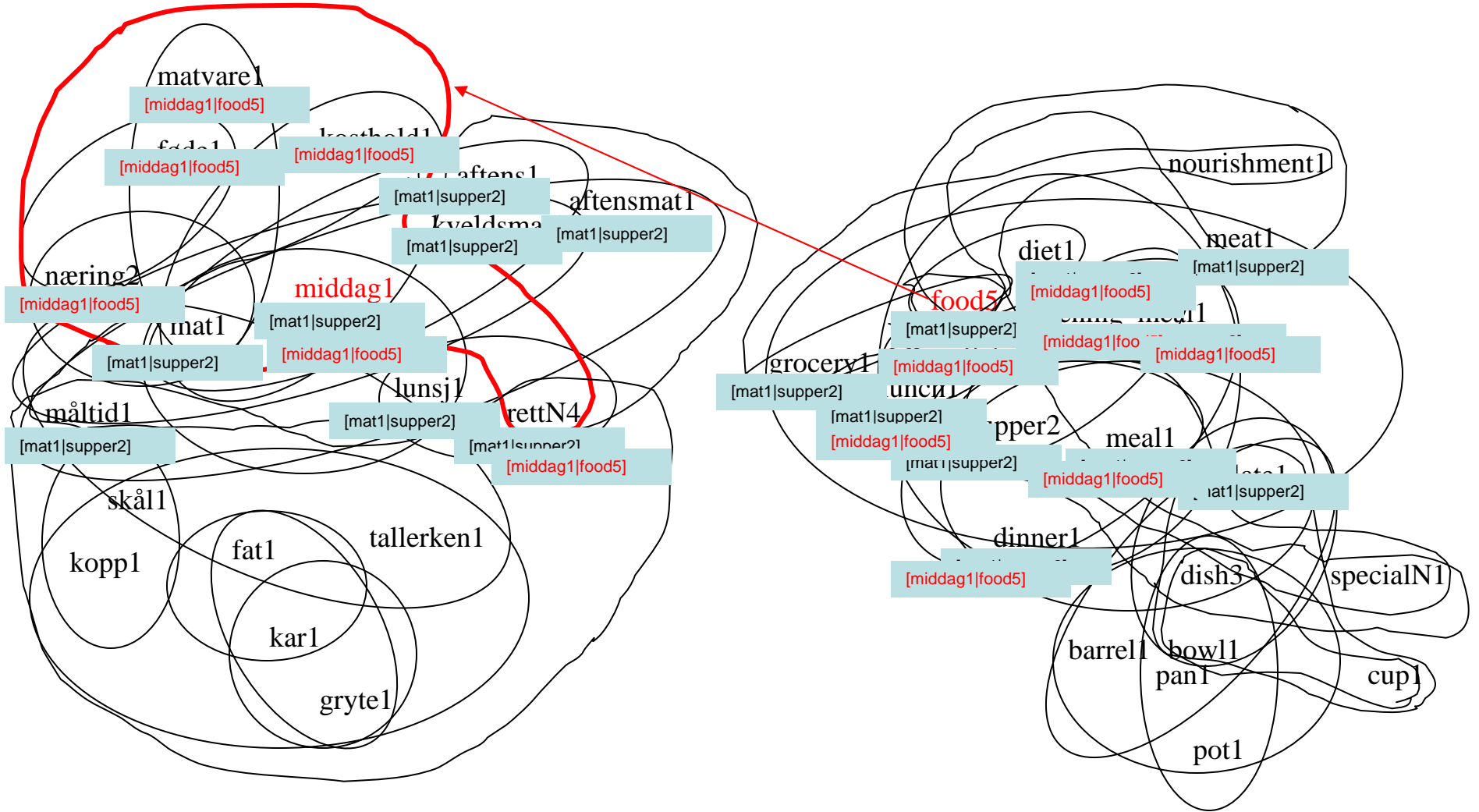
[mat1|supper2]

[middag1|food5]



[mat1|supper2]

[middag1|food5]



<p>aftens1 [mat1 supper2] [aftens1 evening_meal1]</p> <p>aftensmat1 [mat1 supper2] [lunsj1 meal1] [kveldsmat1 meal1] [aftensmat1]</p> <p>fat1 [fat1 dish3]</p> <p>føde1 [middag1 food5] [føde1 grocery1]</p> <p>gryte1 [måltid1 dish3] [fat1 dish3] [kar1 dish3] [gryte1 bowl1]</p> <p>kar1 [kar1 dish3]</p>	<p>kopp1 [måltid1 dish3] [fat1 dish3] [kar1 dish3] [skål1 bowl1] [gryte1 bowl1] [kopp1 cup1]</p> <p>kosthold1 [middag1 food5] [kosthold1 diet1]</p> <p>kveldsmat1 [mat1 supper2] [kveldsmat1 meal1]</p> <p>lunsj1 [mat1 supper2] [lunsj1 meal1]</p> <p>mat1 [mat1 supper2]</p>	<p>matvare1 [middag1 food5] [føde1 grocery1] [matvare1]</p> <p>middag1 [mat1 supper2] [middag1 food5]</p> <p>måltid1 [mat1 supper2] [måltid1 dish3]</p> <p>næring2 [middag1 food5] [næring2 nourishment1] [næring2 meat1]</p>	<p>rettN4 [mat1 supper2] [middag1 food5] [måltid1 dish3] [fat1 dish3] [kar1 dish3] [rettN4 specialN1]</p> <p>skål1 [måltid1 dish3] [fat1 dish3] [kar1 dish3] [skål1 bowl1]</p> <p>tallerken1 [måltid1 dish3] [fat1 dish3] [kar1 dish3] [skål1 bowl1] [gryte1 bowl1] [tallerken1 plate1]</p>
--	---	---	--

<p>*barrel1 [fat1 dish3] [kar1 dish3] [barrel1]</p> <p>*bowl1 [fat1 dish3] [kar1 dish3] [gryte1 bowl1] [skål1 bowl1]</p> <p>*cup1 [skål1 bowl1] [kopp1 cup1]</p> <p>*diet1 [mat1 supper2] [middag1 food5] [kosthold1 diet1]</p>	<p>*dinner1 [mat1 supper2] [middag1 food5] [måltid1 dish3] [kveldsmat1 meal1] [dinner1]</p> <p>dish3 [kar1 dish3] [fat1 dish3] [måltid1 dish3]</p> <p>*evening_meal1 [mat1 supper2] [middag1 food5] [kveldsmat1 meal1] [aftens1 evening_meal1]</p> <p>food5 [mat1 supper2] [middag1 food5]</p> <p>*grocery1 [mat1 supper2] [føde1 grocery1]</p>	<p>*lunch1 [mat1 supper2] [middag1 food5] [lunsj1 meal1] [lunch1]</p> <p>*meal1 [mat1 supper2] [middag1 food5] [måltid1 dish3] [kveldsmat1 meal1] [lunsj1 meal1]</p> <p>*meat1 [mat1 supper2] [næring2 meat1]</p> <p>*nourishment1 [føde1 grocery1] [næring2 nourishment1]</p>	<p>*pan1 [fat1 dish3] [kar1 dish3] [gryte1 bowl1] [pan1]</p> <p>*plate1 [mat1 supper2] [måltid1 dish3] [fat1 dish3] [skål1 bowl1] [tallerken1 plate1]</p> <p>*pot1 [kar1 dish3] [gryte1 bowl1] [pot1]</p> <p>specialN1 [lunsj1 meal1] [rettN4 specialN1]</p> <p>supper2 [mat1 supper2]</p>
---	--	--	---

Assumption:

Feature set inclusion expresses hyperonymy/hyponymy

mat1

[mat1|supper2]



middag1

[mat1|supper2]

[middag1|food5]

food5

[mat1|supper2]

[middag1|food5]



***lunch1**

[mat1|supper2]

[middag1|food5]

[lunsj1|meal1]

[lunch1]

Assumption:

Feature set inclusion expresses hyperonymy/hyponymy

mat1

[mat1|supper2]



middag1

[mat1|supper2]

[middag1|food5]

food5

[mat1|supper2]

[middag1|food5]



***lunch1**

[mat1|supper2]

[middag1|food5]

[lunsj1|meal1]

[lunch1]

The set of senses in a field can be seen as a *partially ordered set*, ordered by set inclusion.

A *semilattice* (upper):

A partially ordered set in which each pair of elements has a *least upper bound*.

A *semilattice* (upper):

A partially ordered set in which each pair of elements has a *least upper bound*.

Applied to our case:

For each pair of feature sets,

- either one set includes the other,
- or there is a third element consisting of the intersection of the two sets.

A *semilattice* (upper):

A partially ordered set in which each pair of elements has a *least upper bound*.

Applied to our case:

For each pair of feature sets,

- either one set includes the other,
- or there is a third element consisting of the intersection of the two sets.

Constructing a semilattice allows us to compare the distances between all senses

Constructing a semilattice:

If two feature sets intersect without inclusion,

busy2

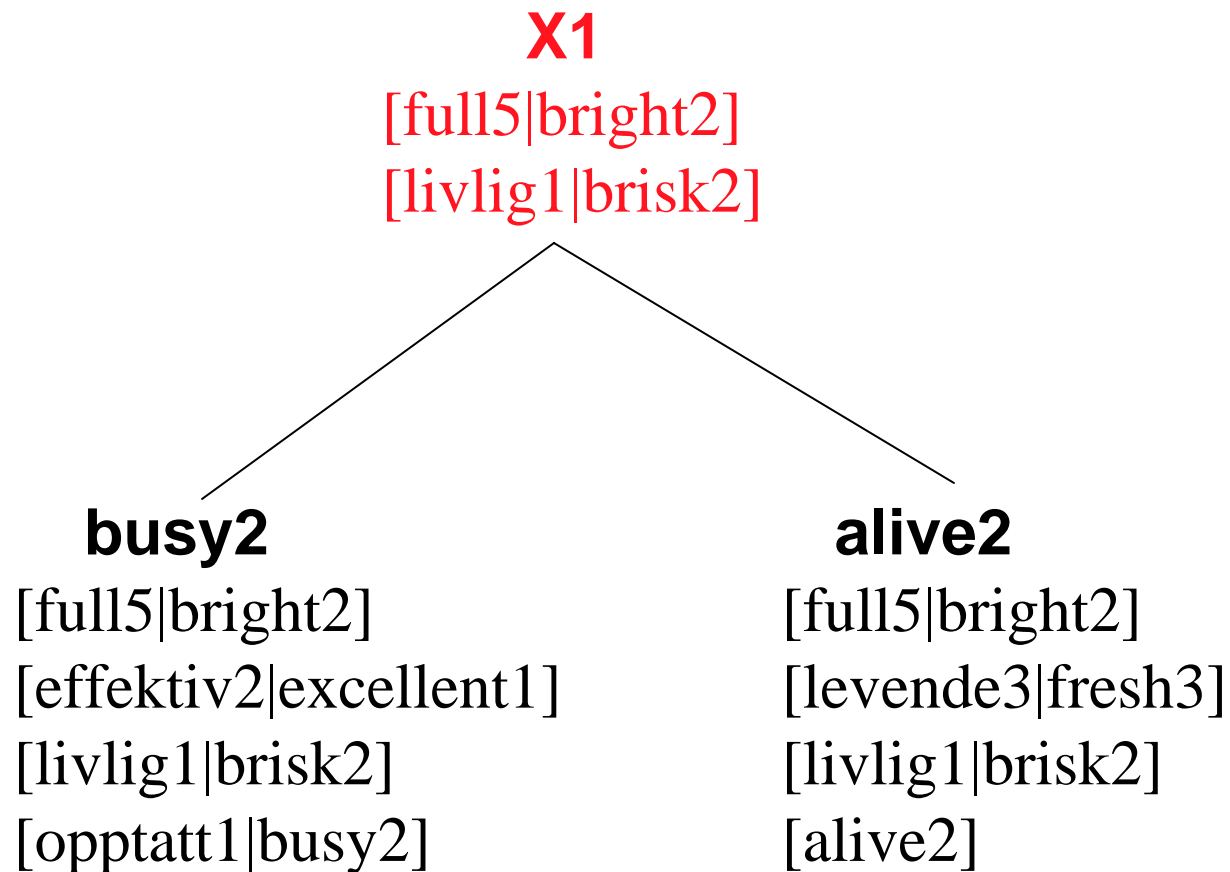
[full5|bright2]
[effektiv2|excellent1]
[livlig1|brisk2]
[opptatt1|busy2]

alive2

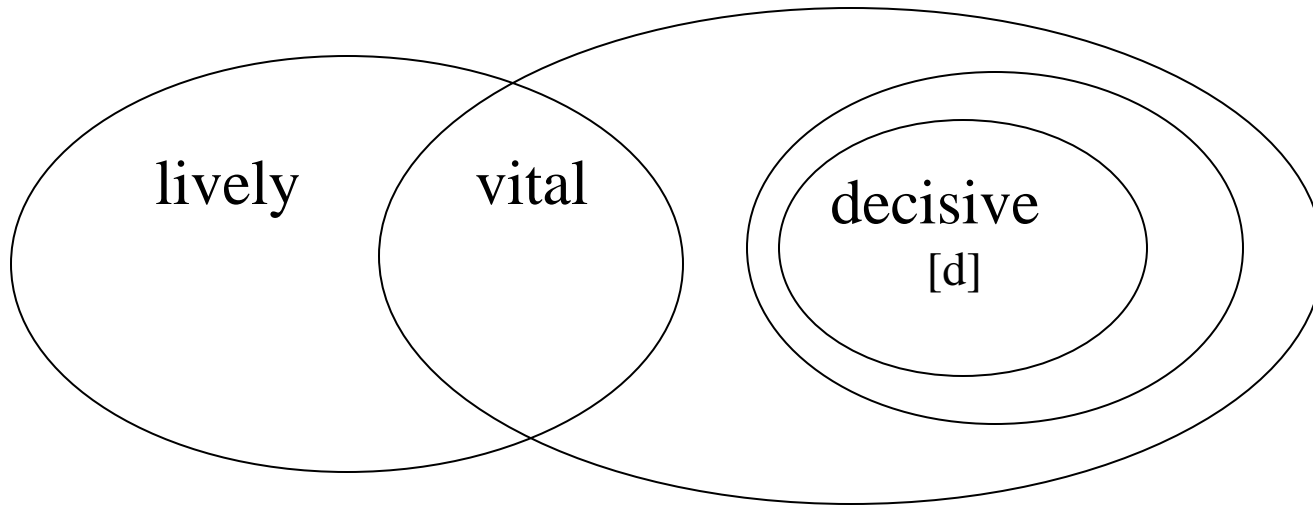
[full5|bright2]
[levende3|fresh3]
[livlig1|brisk2]
[alive2]

Constructing a semilattice:

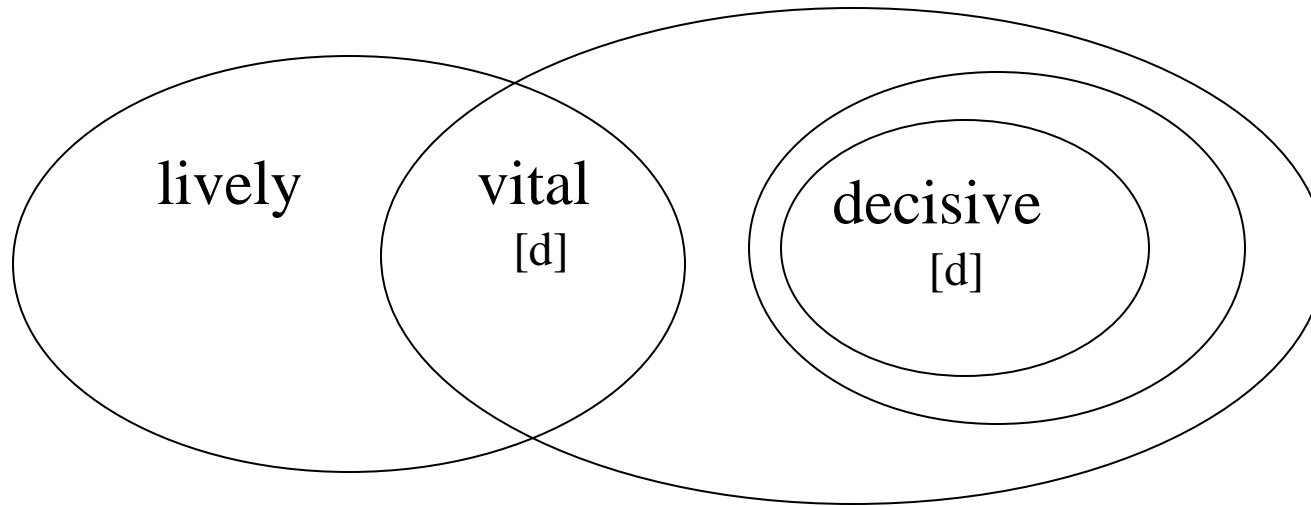
If two feature sets intersect without inclusion, a «virtual» mother node with the intersection of the two feature sets is constructed:



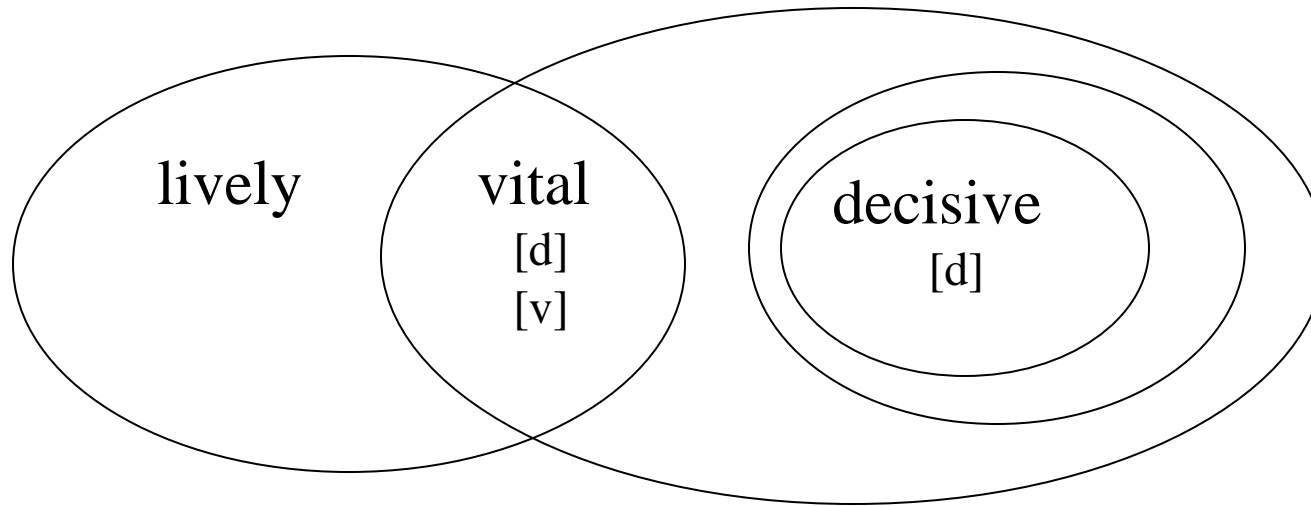
A sense only transfers its *own* features downwards,
not its *inherited* features:



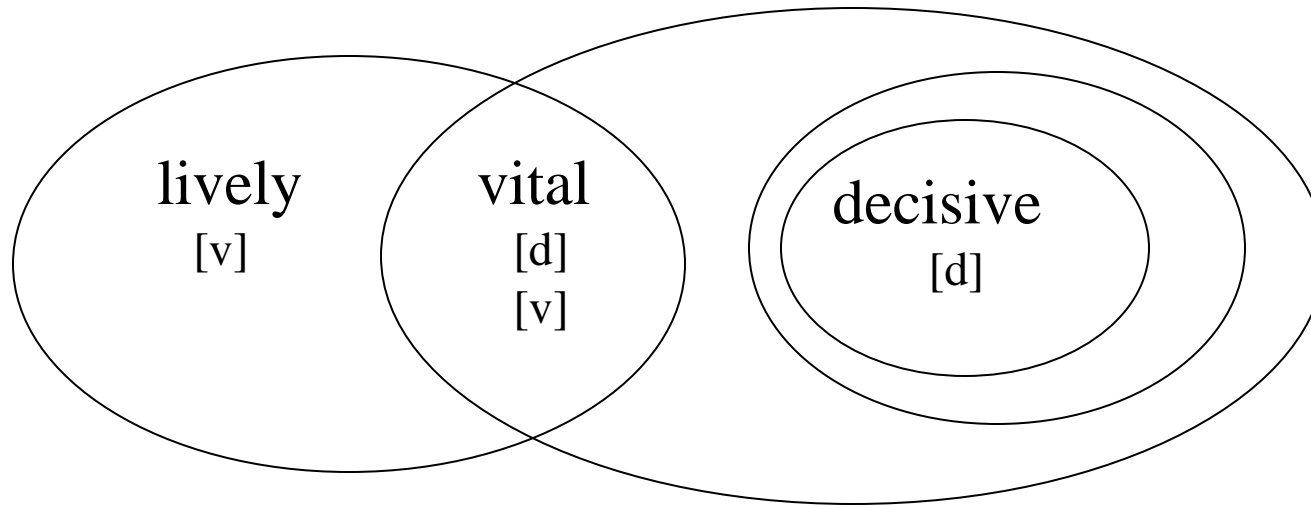
A sense only transfers its *own* features downwards,
not its *inherited* features:



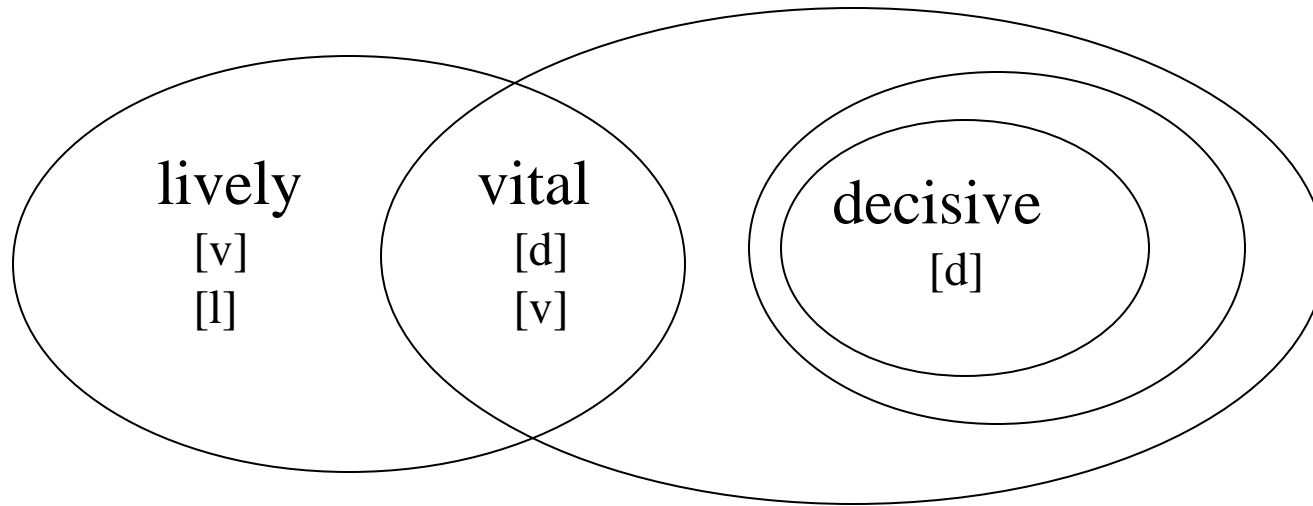
A sense only transfers its *own* features downwards,
not its *inherited* features:



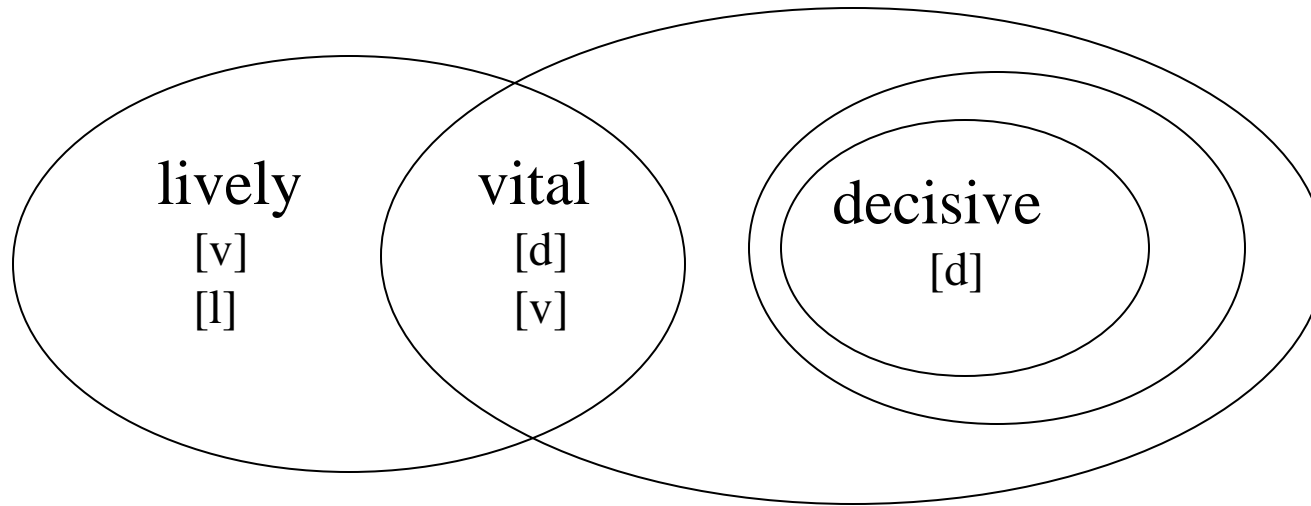
A sense only transfers its *own* features downwards,
not its *inherited* features:



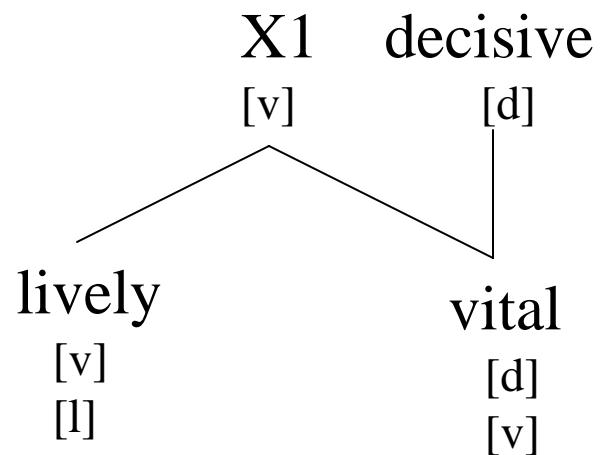
A sense only transfers its *own* features downwards,
not its *inherited* features:



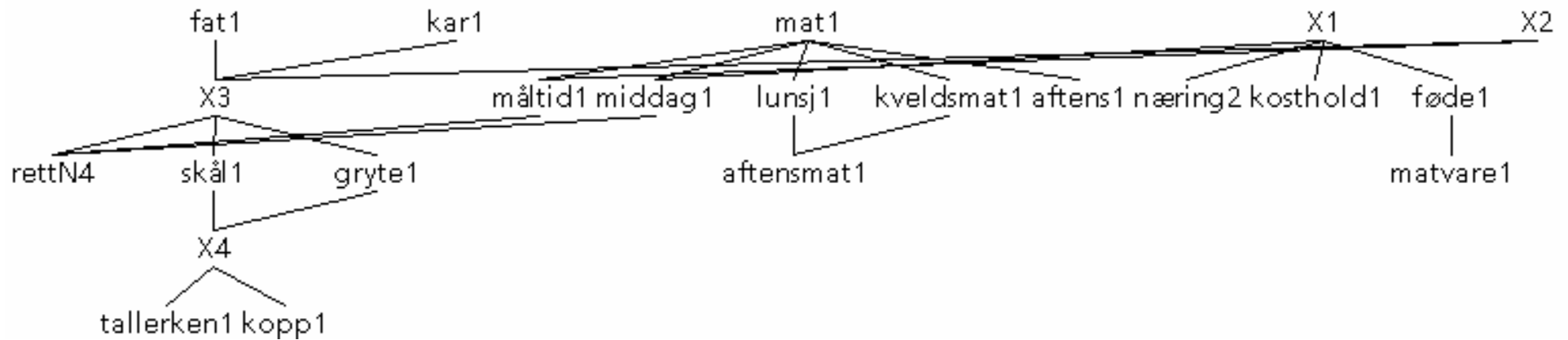
A sense only transfers its *own* features downwards,
not its *inherited* features:



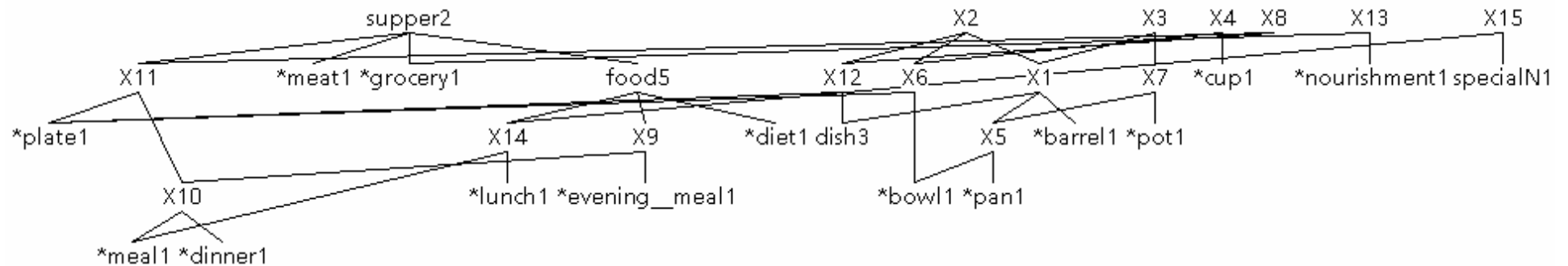
Resulting lattice:



Norwegian semantic field as lattice graph



Corresponding English semantic field



Features assigned to *sweet1*:

sweet1

[god3|good1]

[fin2|nice2]

[pen1|gentle3]

[vakker1|soft2]

[snill1|pleasant1]

[deilig1|splendid3]

[frisk4|sweet1]

[blid3|sweet1]

The feature denotations:

[god3|good1]:

(able1 accurate1 adept1 adequate2 affectionate1 all_right2 amiable2 appropriate5 attractive4 beautiful2 beneficial1 benign3 bright2 burning3 charming2 clean1 clear1 close3 comfortable2 comforting3 competent2 confident2 correct1 cozy2 cute1 decent2 delicious1 delightful2 detailed3 dishy1 easy1 efficient2 elegant3 excellent2 fair2 fancy1 favourable1 fine1 firmA1 first-class3 first-rate2 fit3 fortunate1 fresh3 friendly2 full2 genuine2 good1 handsome2 happy3 healthy2 high3 hot2 joyful2 kind1 kindly1 long3 lovely2 lucky2 magnificent3 marvellous1 neat2 nice2 okay1 peaceful1 perfect3 placid2 pleasant1 pleased2 pleasing1 pleasurable1 plentiful1 plenty1 polite2 positive1 pretty2 proficient1 quite_certain1 real2 reassuring2 respectable3 right2 ripe1 safe2 satisfactory1 satisfying1 secure2 sizeable1 smart2 smooth3 soft2 solid2 sound2 spectacular2 steady1 strong3 successful2 suited1 superb2 superior5 sure1 sweet1 talented2 thorough1 tidy1 well2 whole2 wholesome1 wonderful3 worthy2)

[fin2|nice2]:

(attractive4 beautiful2 breathtaking2 charming2 comfortable2 cute1 delicate3 dishy1 easy1 elegant3 enchanting1 excellent2 fancy1 fine1 first-class3 gentle3 glorious4 graceful2 handsome2 impressive2 lovely2 magnificent3 marvellous1 neat2 nice2 okay1 perfect3 pleasurable1 polite2 pretty2 pure2 slight3 smart2 soft2 splendid3 sweet1 thin2 wonderful3)

[pen1|gentle3]:

(attractive4 beautiful2 charming2 clean1 cute1 dishy1 elegant3 enchanting1 fancy1 fine1 first-class3 formal1 gentle3 graceful2 handsome2 lovely2 neat2 pleasant1 polite2 pretty2 soft2 sweet1 tidy1)

[vakker1|soft2]:

(attractive⁴ charming² cute¹ delightful² dshy¹ enchanting¹ fair² fancy¹ graceful² handsome² lovely² magnificent³ mild² ornate² pleasant¹ pleasurable¹ pretty² soft² sweet¹)

[snill1|pleasant1]:

(all_right² amiable² benign³ friendly² good-humoured¹ good-natured³ jolly¹ kind¹ kindly¹ mild³ pleasant¹ pleasing¹ polite² smiling² sweet¹)

[deilig1|splendid3]:

(beautiful² charming² cute¹ enchanting¹ delicious¹ delightful² pleasureable¹ splendid³ sweet¹)

[frisk4|sweet1]:

(all_right² brisk⁵ eager² fit³ fresh³ healthy² new¹ pert² sweet¹ well²)

[blid3|sweet1]:

(amiable² blithe³ cheerful⁴ cheery¹ good-humoured¹ good-natured³ jolly¹ kind¹ kindly¹ merry¹ mild³ smiling² sweet¹)

SynsetLimit (integer)

Given the feature set F of a sense s :

SynsetLimit (integer)

Given the feature set F of a sense s :

For a feature $f \in F$, if the cardinality of f 's denotation exceeds *SynsetLimit*, then its members are not included among possible synonyms or related words (except if denoted by other features).

SynsetLimit (integer)

Given the feature set F of a sense s :

For a feature $f \in F$, if the cardinality of f 's denotation exceeds *SynsetLimit*, then its members are not included among possible synonyms or related words (except if denoted by other features).

The sense from which such a feature was constructed is considered a *hyperonym* of s .

Synonyms (close) and *related words* (less close):

Defined in terms of feature sharing
(number and kind of shared features)

Related subsenses:

The division of a given sense into related subsenses does not have a unique solution.

Related subsenses:

The division of a given sense into related subsenses does not have a unique solution.

We may want to vary the granularity of such divisions according to our purposes.

Related subsenses:

The division of a given sense into related subsenses does not have a unique solution.

We may want to vary the granularity of such divisions according to our purposes.

The parameter *OverlapThreshold* allows us to do this.

Related subsenses:

The division of a given sense into related subsenses does not have a unique solution.

We may want to vary the granularity of such divisions according to our purposes.

The parameter *OverlapThreshold* allows us to do this.

OverlapThreshold [0,1]:

Specifies the minimal *overlap degree* of two feature denotations for the features to be counted as belonging to the same subsense.

Intersecting feature denotations:

[frisk4|sweet1]

all_right2 brisk5
eager2 fit3 fresh3 healthy2
new1 pert2 well2

[blid3|sweet1]

blithe3 cheerful4 cheery1 merry1
amiable2 good-humoured1
good-natured3 jolly1 kind1
kindly1 mild3 smiling2

[snill1|pleasant1]

all_right2benign3 friendly2
pleasing1 polite2
pleasant1

[vakker1|soft2]

attractive4 delightful2
dishy1 fair2 fancy1 graceful2
handsome2 lovely2 magnificent3 mild2
ornate2 pretty2 soft2
charming2 cute1
enchancing1 pleasurable1

[deilig1|splendid3]

beautiful2 delicious1 delightful2
splendid3

Web demo...

<http://ling.uib.no/~helge/mirrwebguide.html>

god

hyggelig

flink

skarp

snill

yndig

pen

skjønn

velsmakende

good

nice

clever

sharp

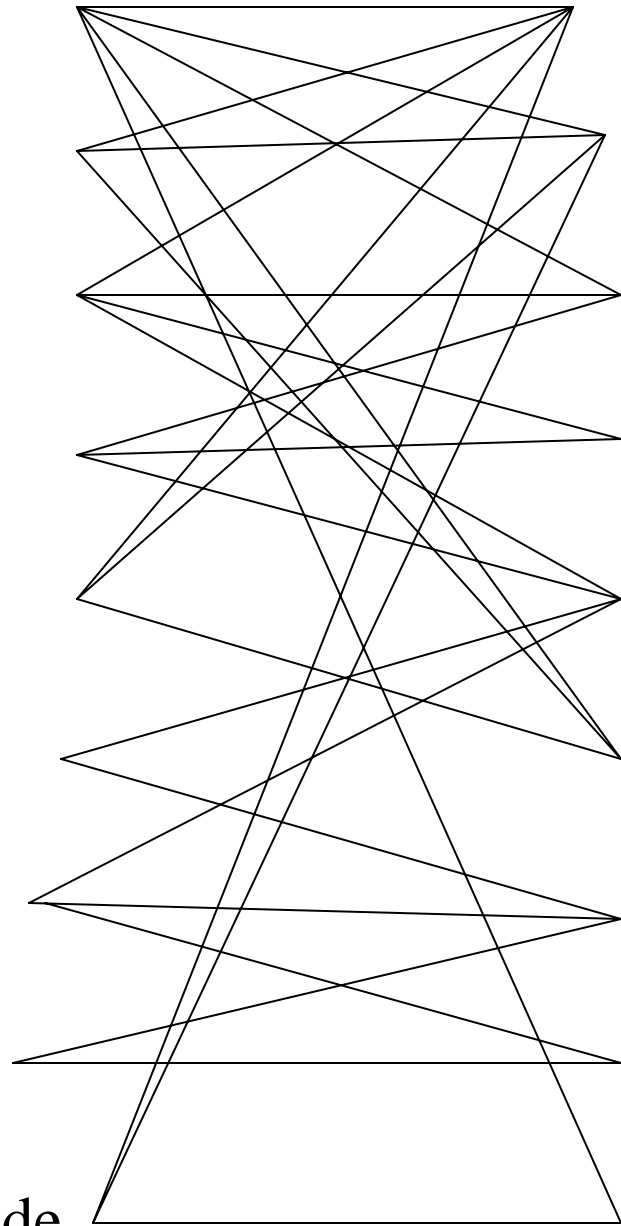
cute

kind

pretty

beautiful

delicious



god

hyggelig

flink

skarp

snill

yndig

pen

skjønn

velsmakende

good

nice

clever

sharp

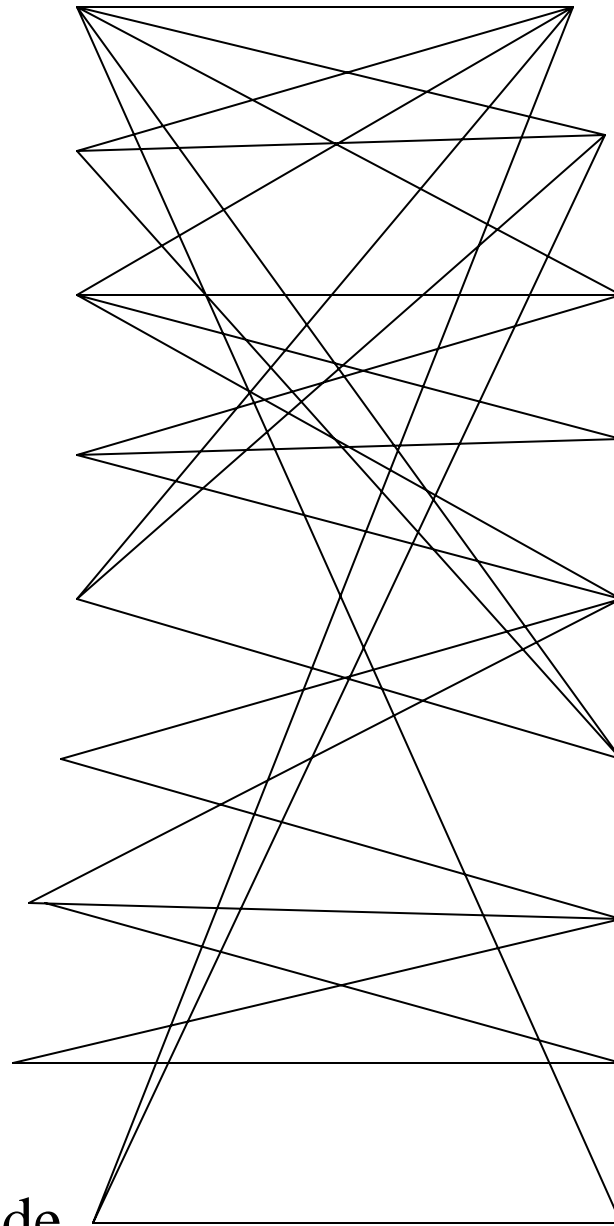
cute

kind

pretty

beautiful

delicious



The first *t*-image of 'good'

god

hyggelig

flink

skarp

snill

yndig

pen

skjønn

velsmakende

good

nice

clever

sharp

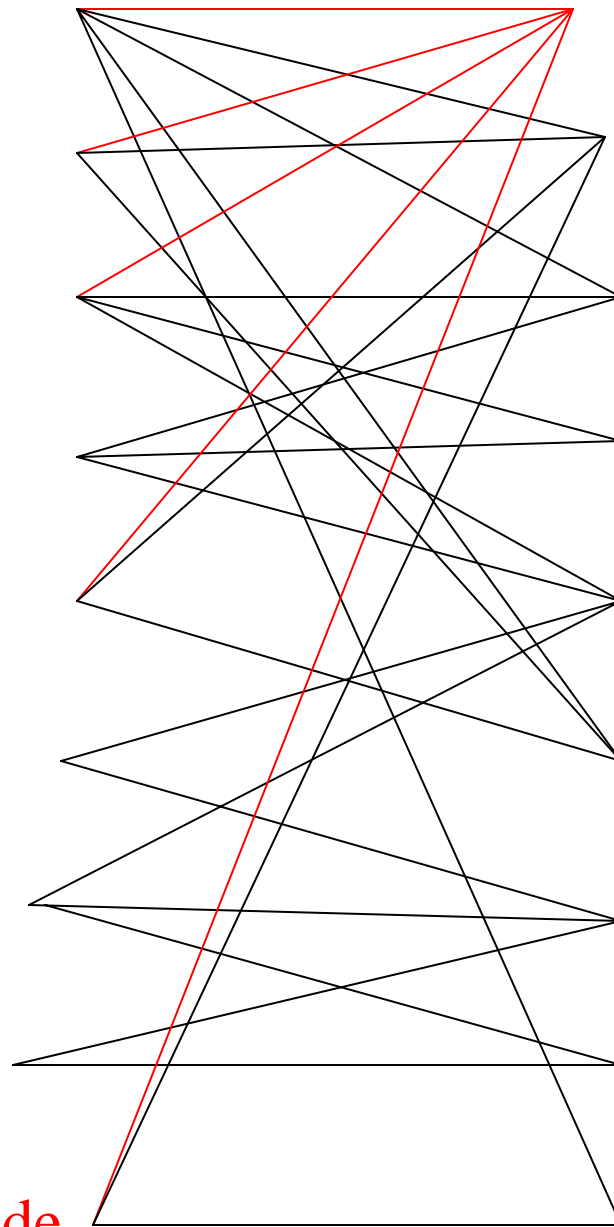
cute

kind

pretty

beautiful

delicious



god

hyggelig

flink

skarp

snill

yndig

pen

skjønn

velsmakende

good

nice

clever

sharp

cute

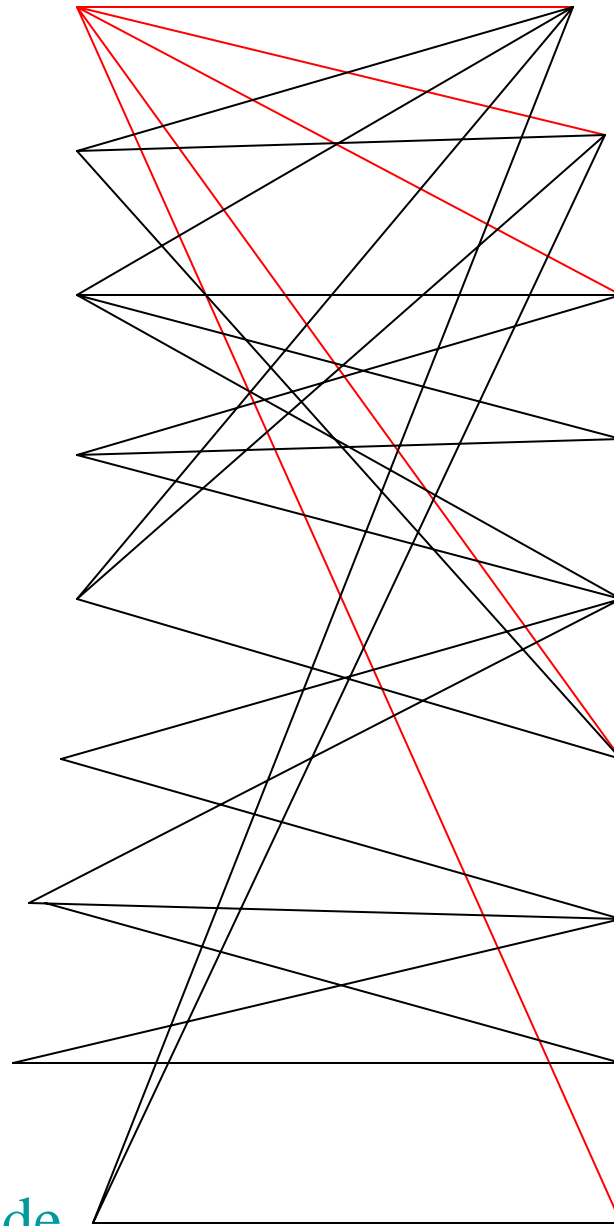
kind

pretty

beautiful

delicious

The
inverse
t-image
of
'good'
I



god

hyggelig

flink

skarp

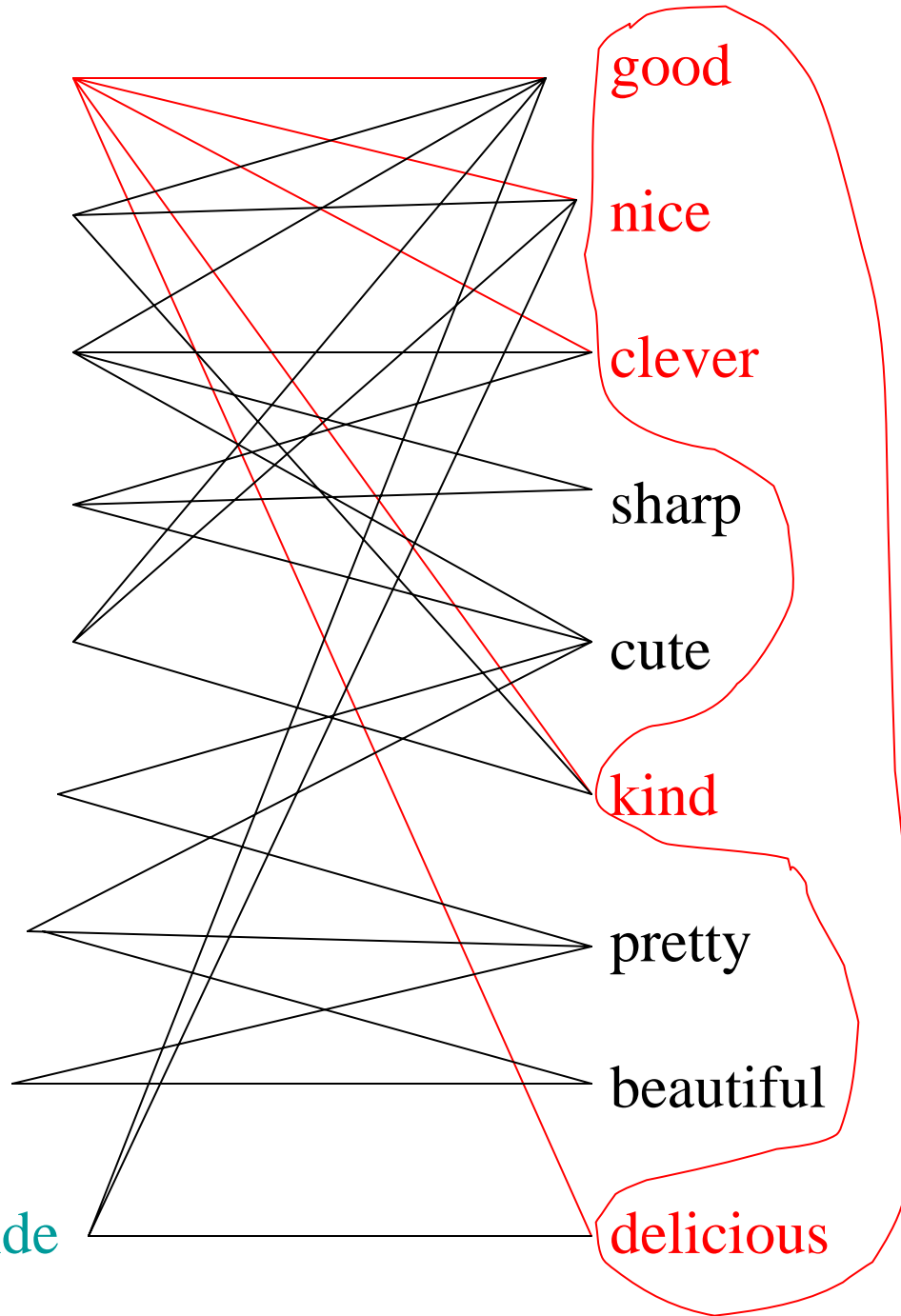
snill

yndig

pen

skjønn

velsmakende



good

nice

clever

sharp

cute

kind

pretty

beautiful

delicious

The
inverse
t-image
of
'good'
I

god

hyggelig

flink

skarp

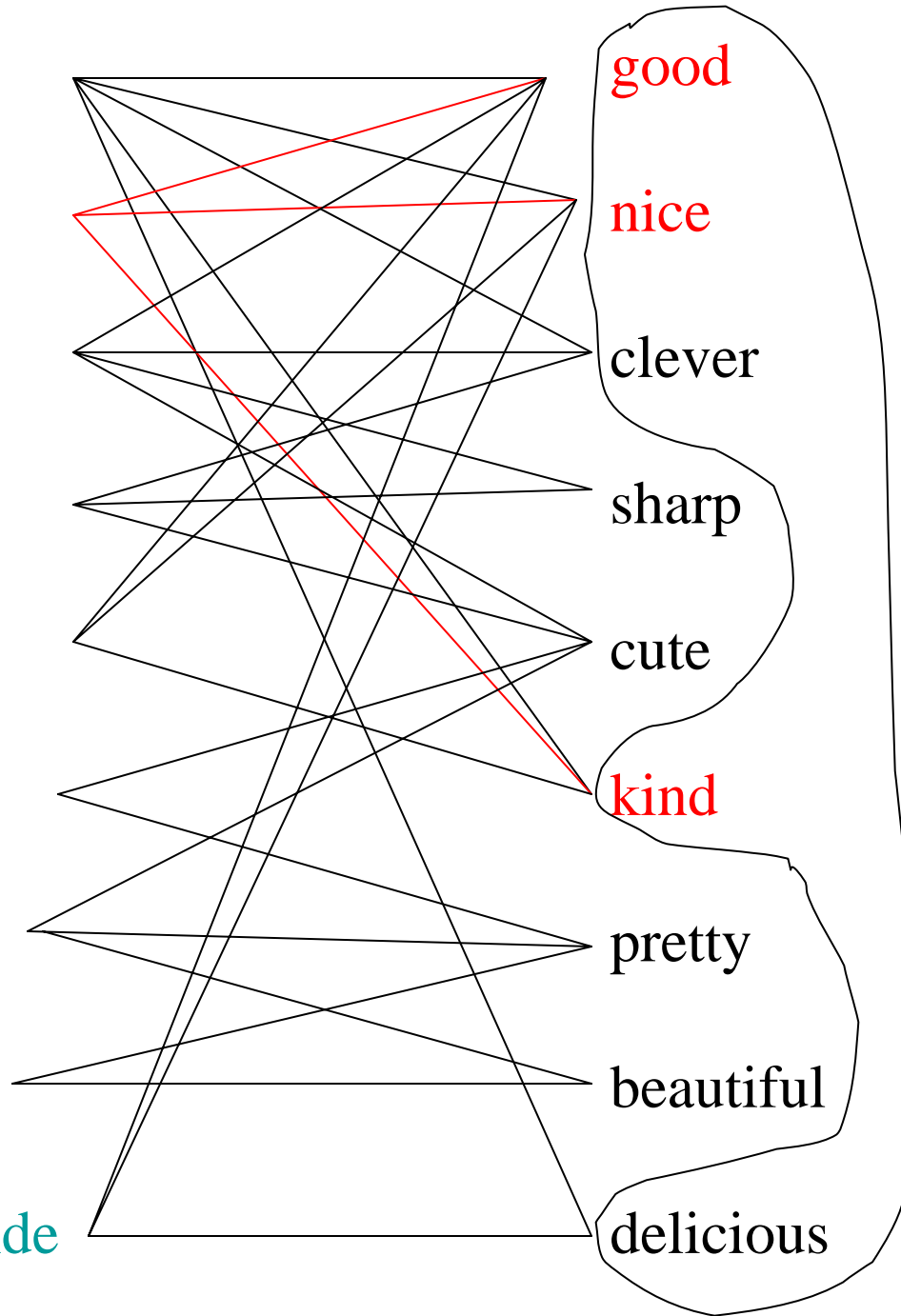
snill

yndig

pen

skjønn

velsmakende



good

nice

clever

sharp

cute

kind

pretty

beautiful

delicious

The
inverse
t-image
of
'good'
II

god

hyggelig

flink

skarp

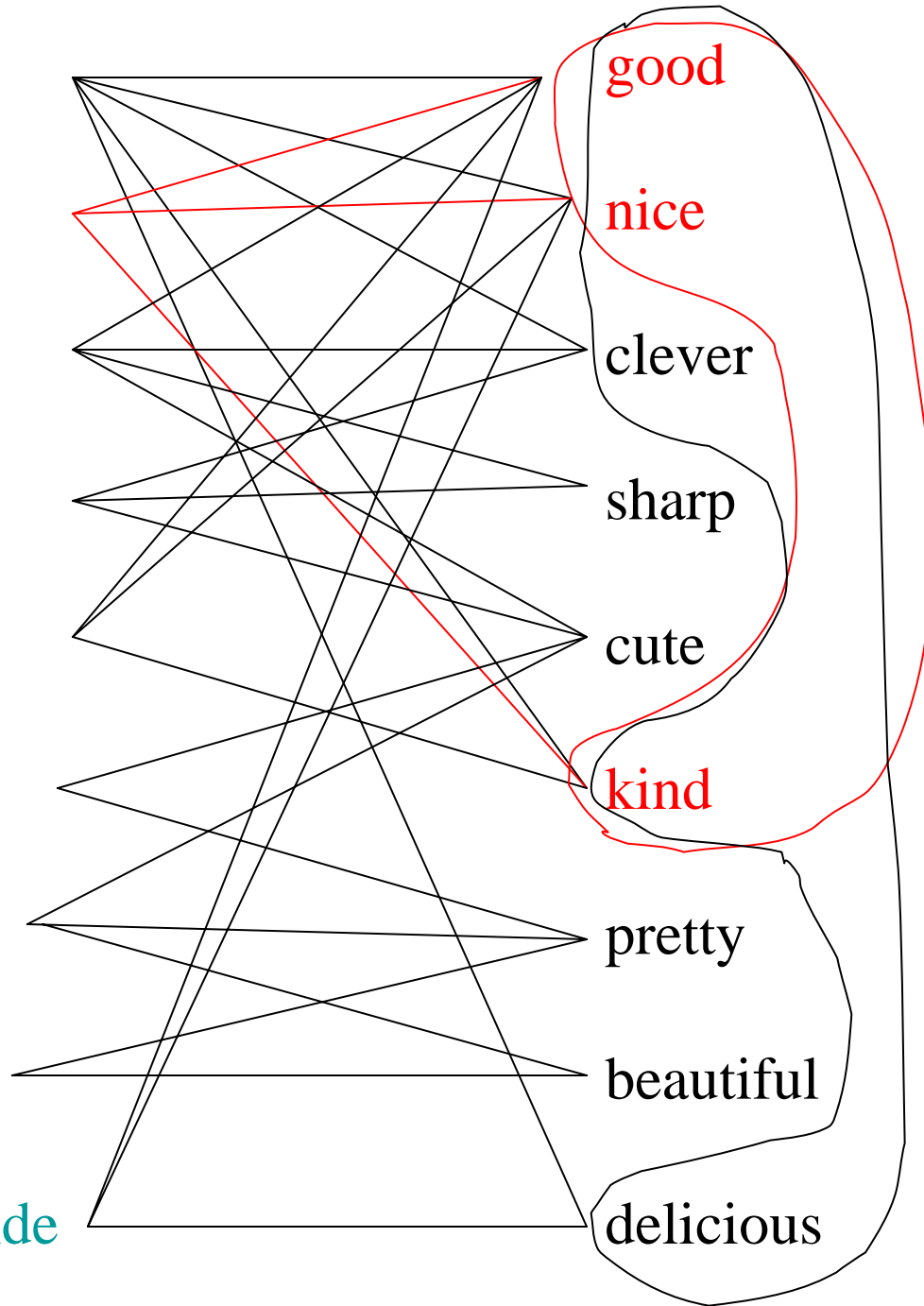
snill

yndig

pen

skjønn

velsmakende



good

nice

clever

sharp

cute

kind

pretty

beautiful

delicious

The
inverse
t-image
of
'good'
II

god

hyggelig

flink

skarp

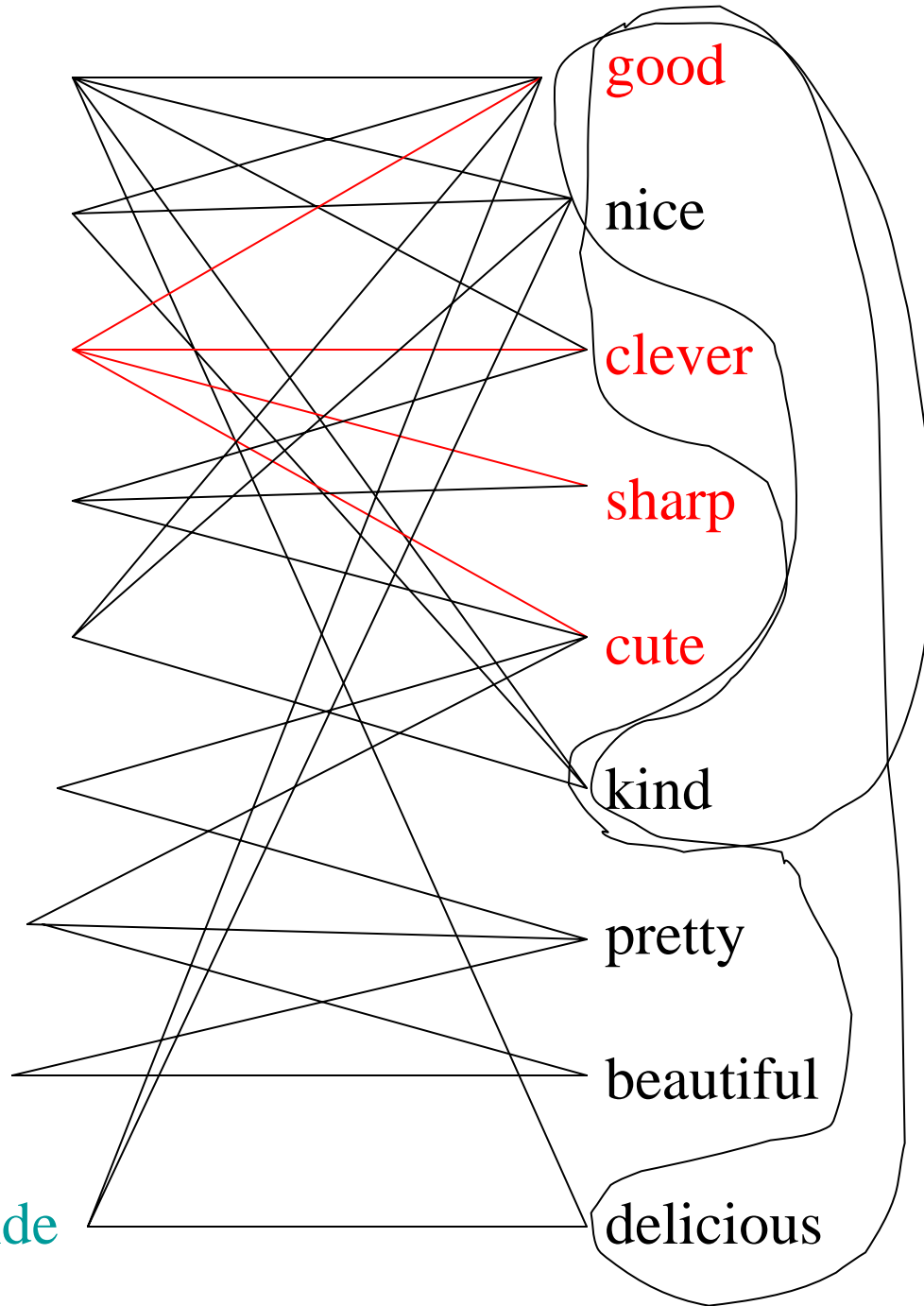
snill

yndig

pen

skjønn

velsmakende



The
inverse
t-image
of
'good'
III

god

hyggelig

flink

skarp

snill

yndig

pen

skjønn

velsmakende

good

nice

clever

sharp

cute

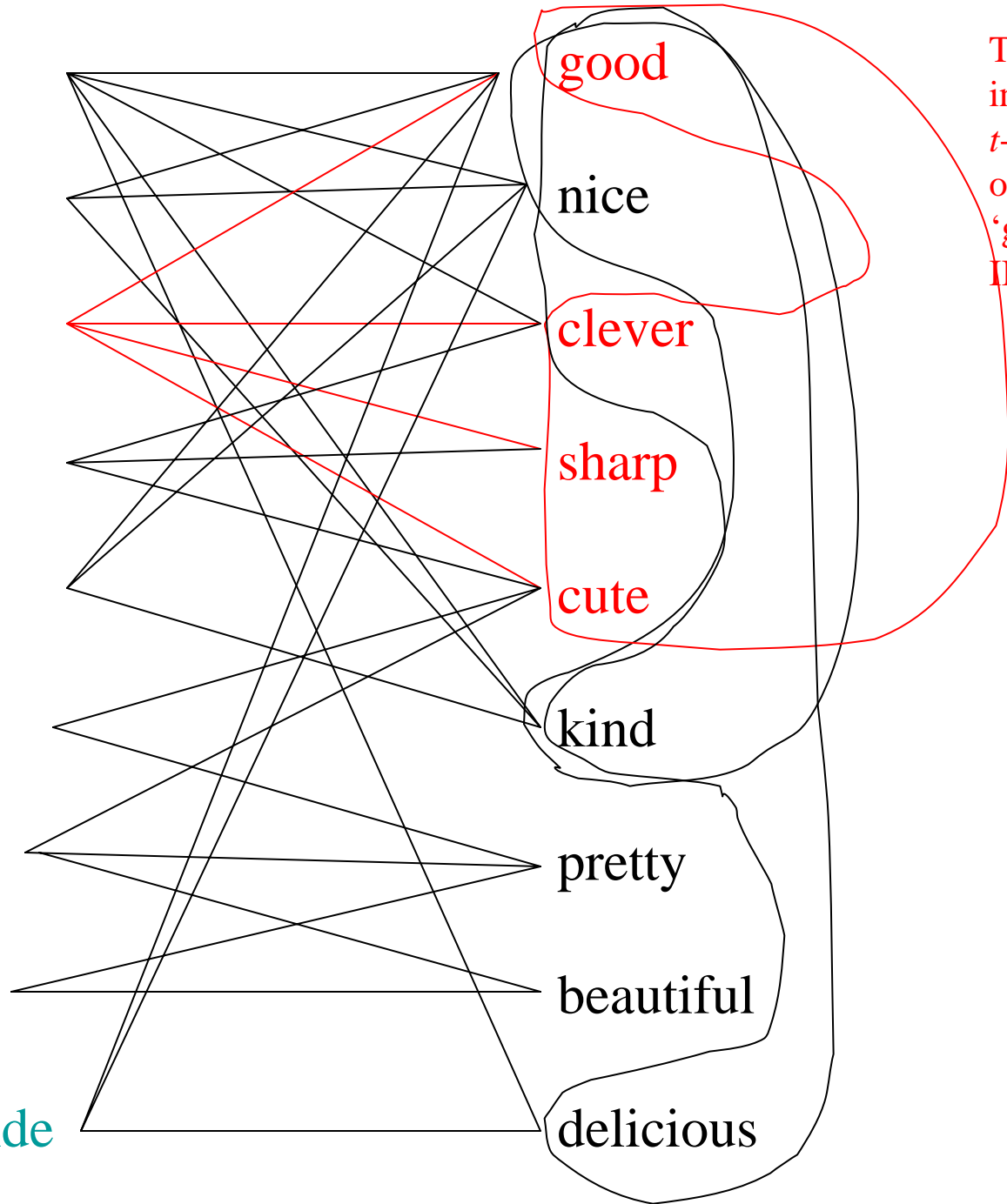
kind

pretty

beautiful

delicious

The
inverse
t-image
of
'good'
III



god

hyggelig

flink

skarp

snill

yndig

pen

skjønn

velsmakende

good

nice

clever

sharp

cute

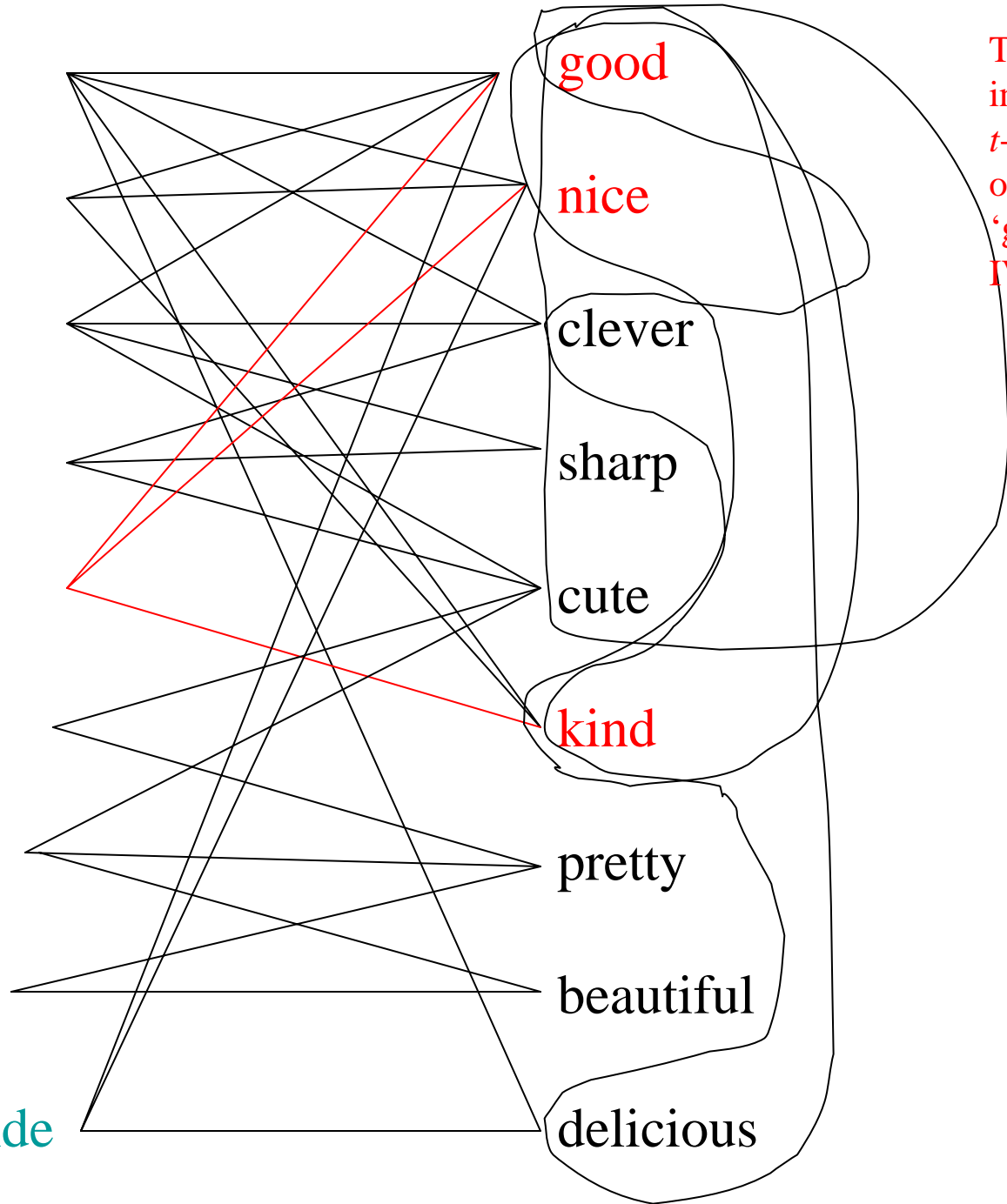
kind

pretty

beautiful

delicious

The
inverse
t-image
of
'good'
IV



god

hyggelig

flink

skarp

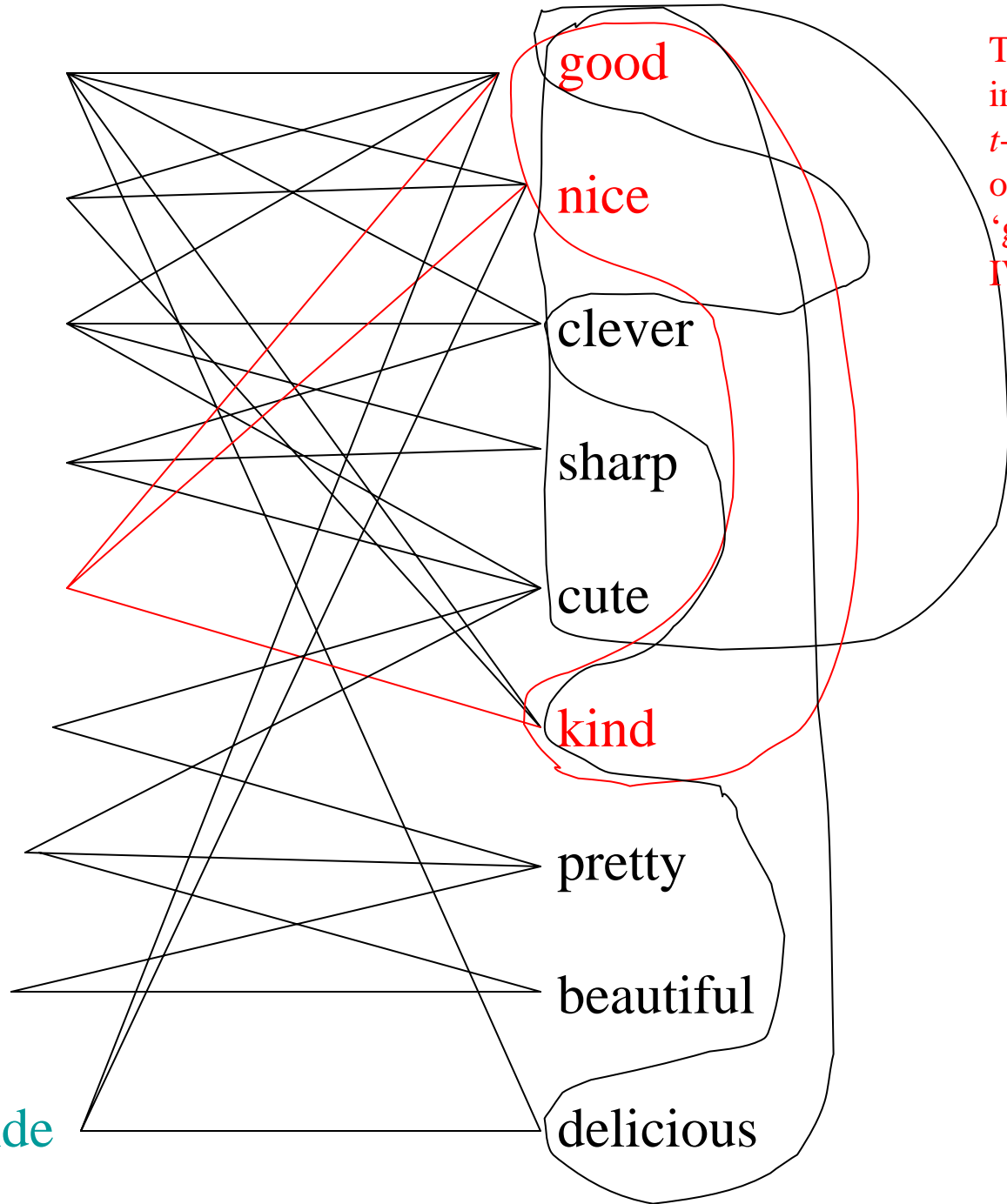
snill

yndig

pen

skjønn

velsmakende



The
inverse
 t -image
of
'good'
IV

god

hyggelig

flink

skarp

snill

yndig

pen

skjønn

velsmakende

good

nice

clever

sharp

cute

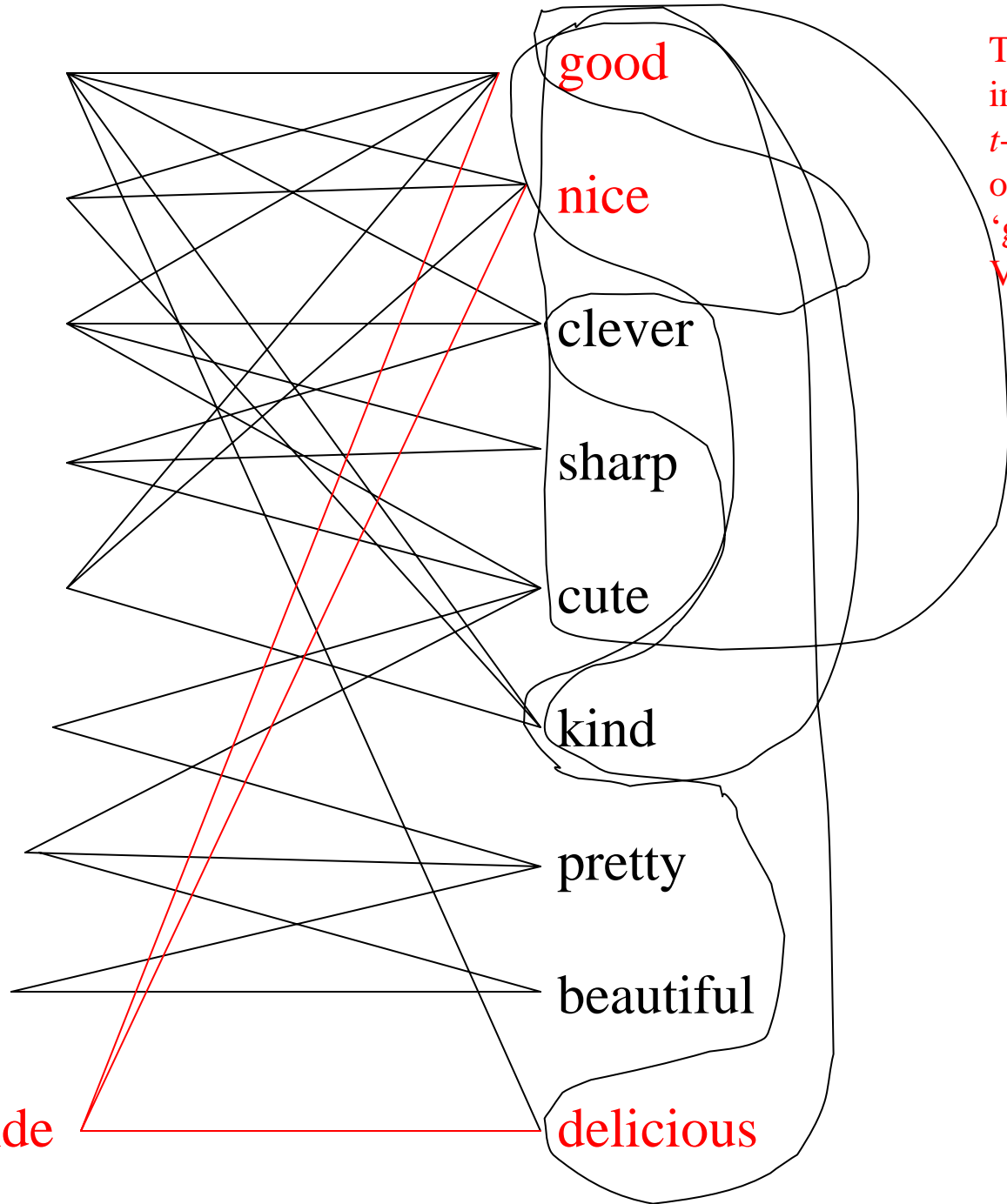
kind

pretty

beautiful

delicious

The
inverse
t-image
of
'good'
 V



god

hyggelig

flink

skarp

snill

yndig

pen

skjønn

velsmakende

good

nice

clever

sharp

cute

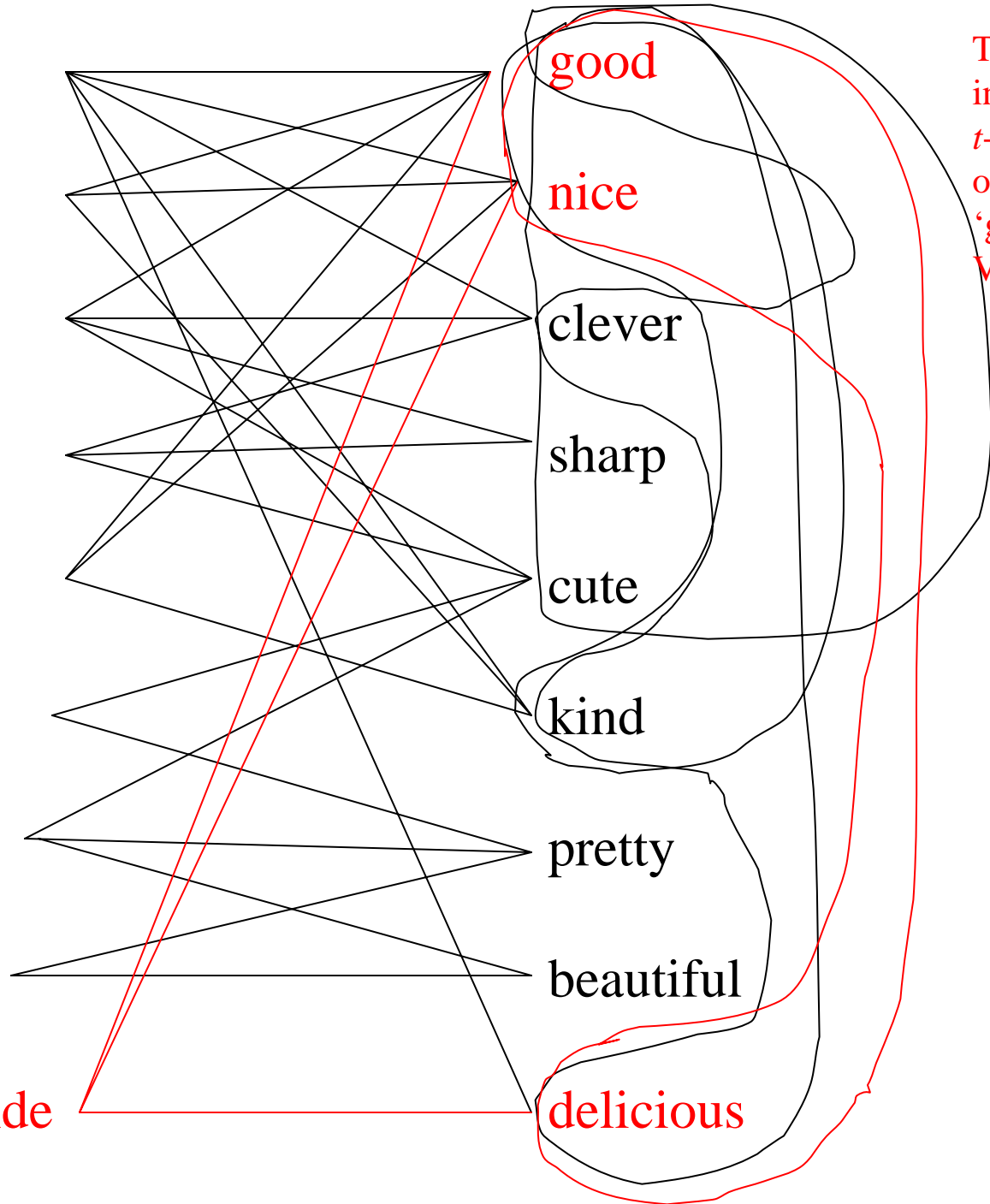
kind

pretty

beautiful

delicious

The
inverse
t-image
of
'good'
 V



god

hyggelig

flink

skarp

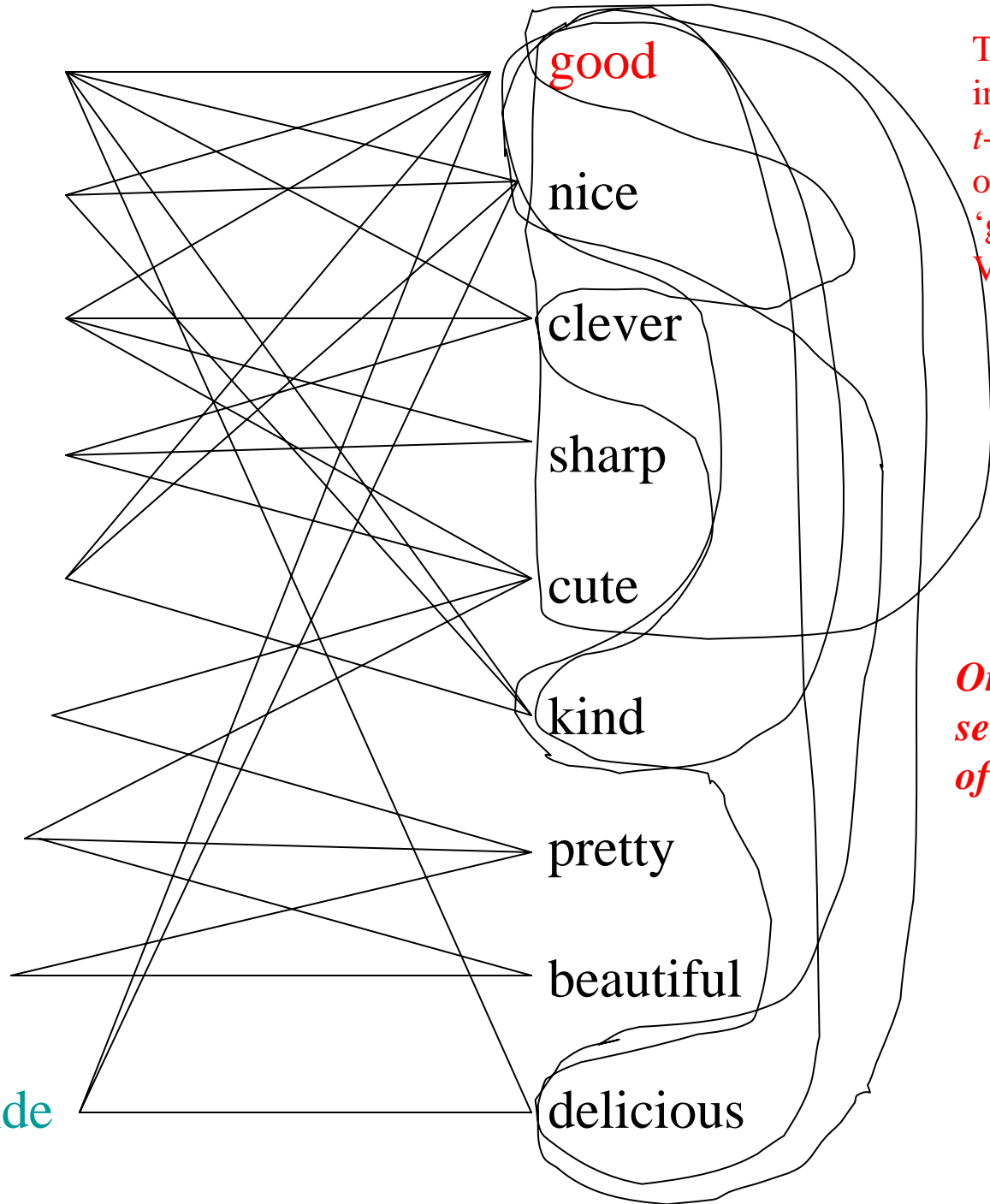
snill

yndig

pen

skjønn

velsmakende



The
inverse
t-image
of
'good'
 V

*One
sense
of 'good'*

god

hyggelig

flink

skarp

snill

yndig

pen

skjønn

velsmakende

good

nice

clever

sharp

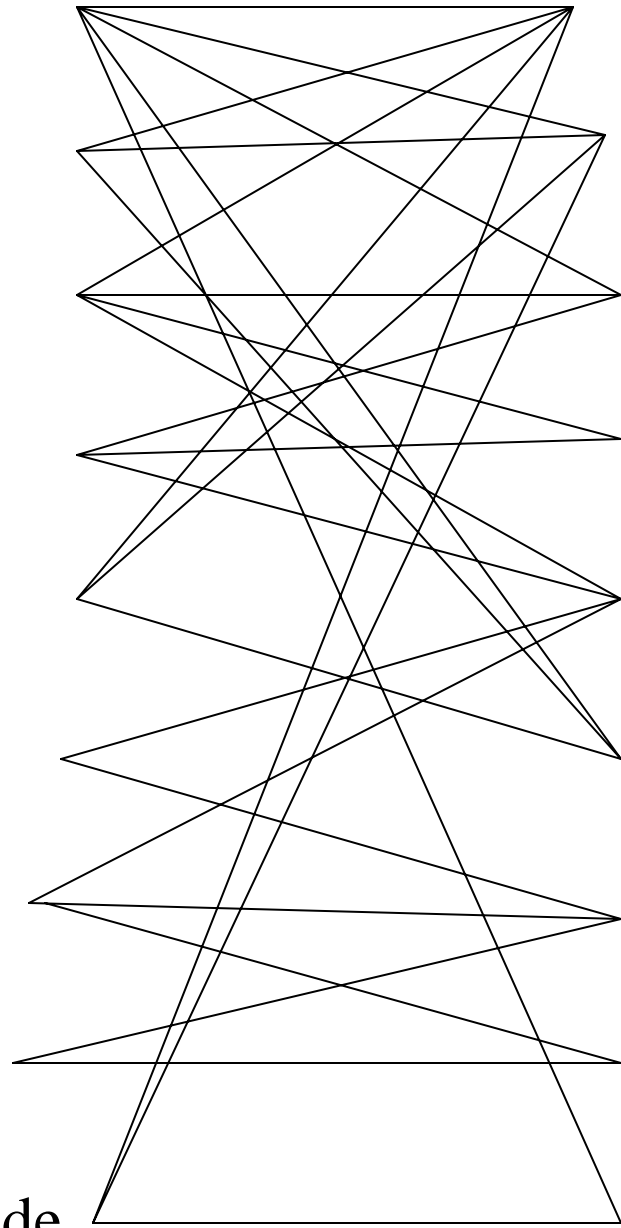
cute

kind

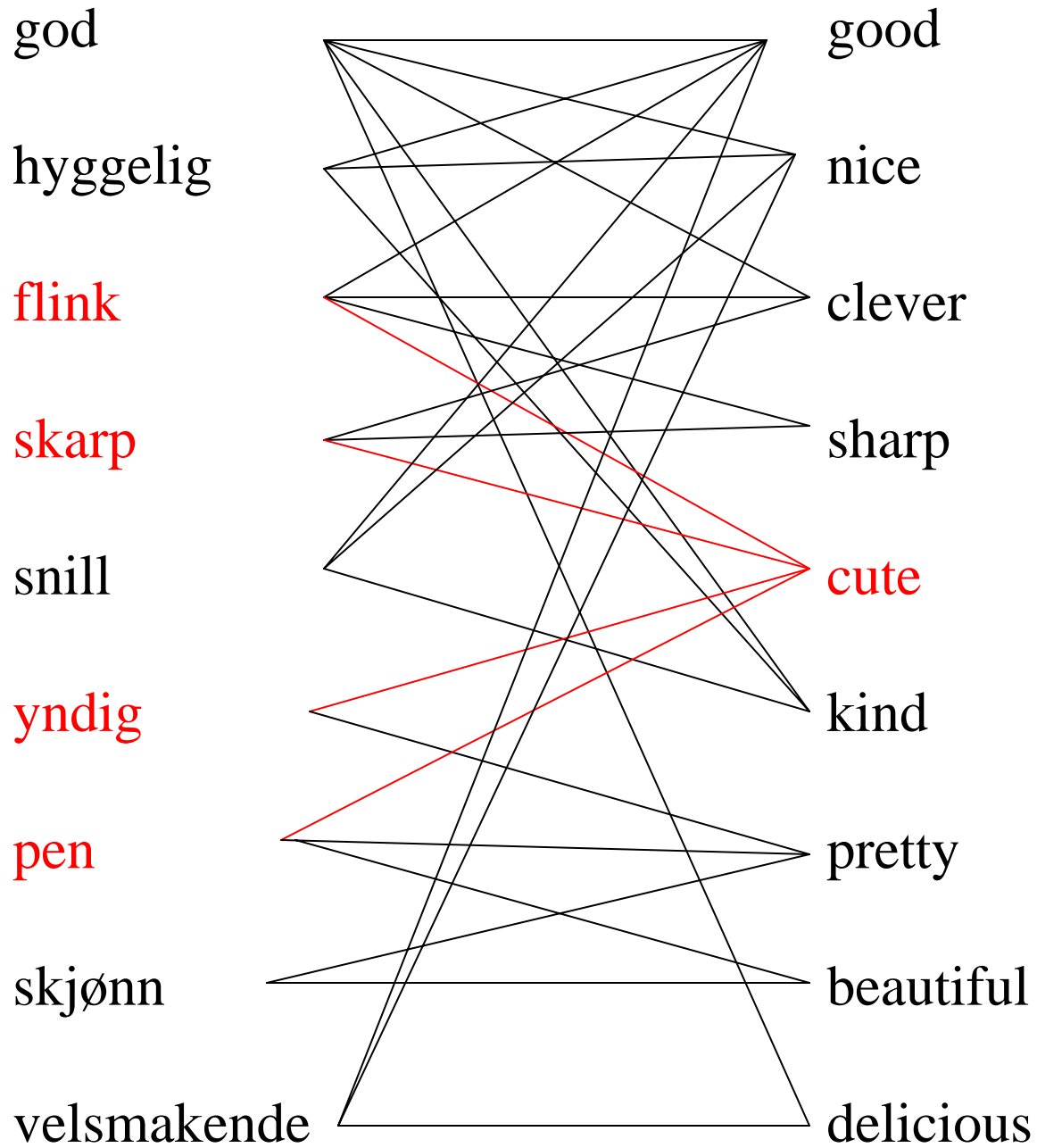
pretty

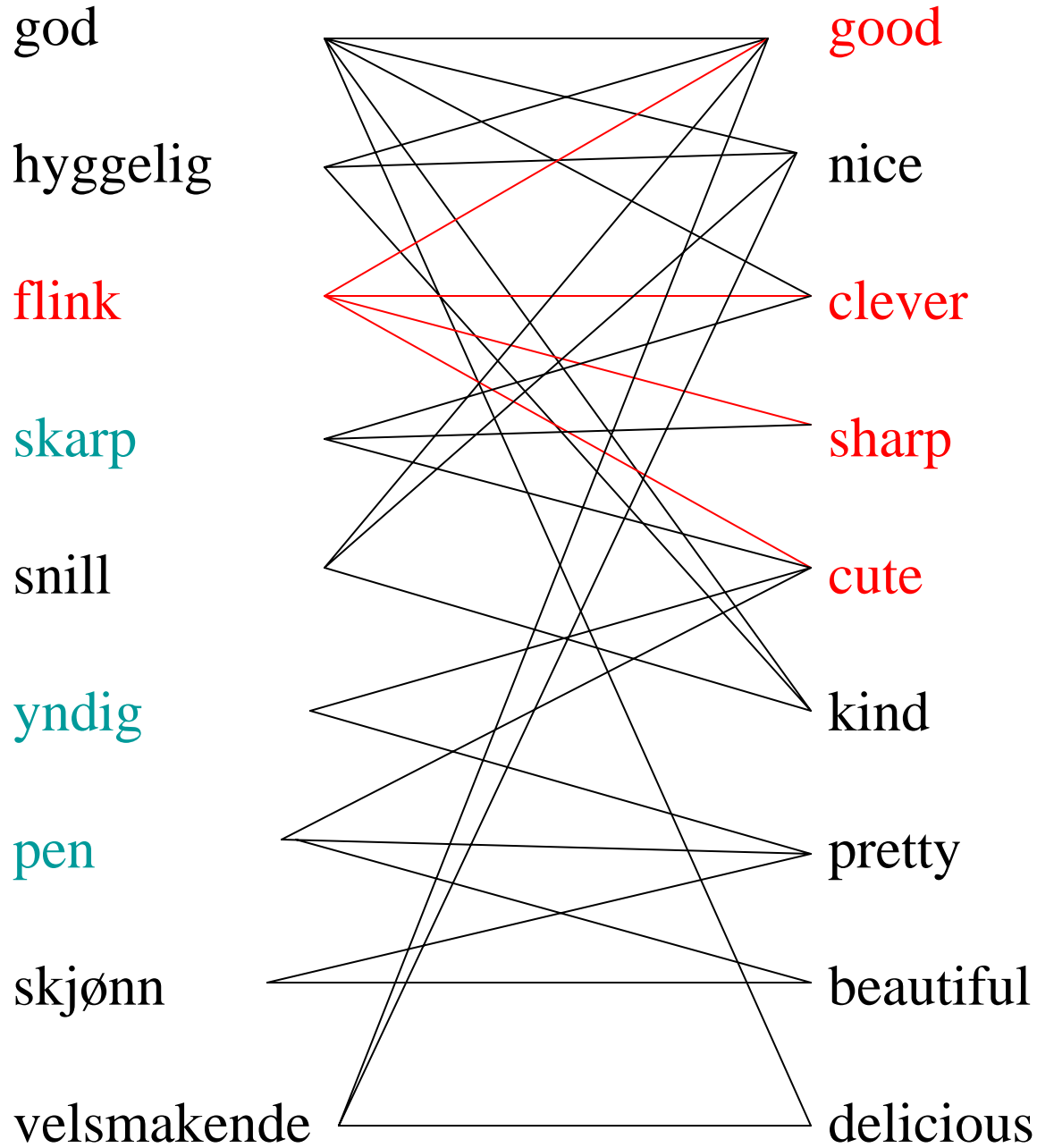
beautiful

delicious



The first *t*-image of 'cute'





The
inverse
t-image
of
'cute'
I

god

hyggelig

flink

skarp

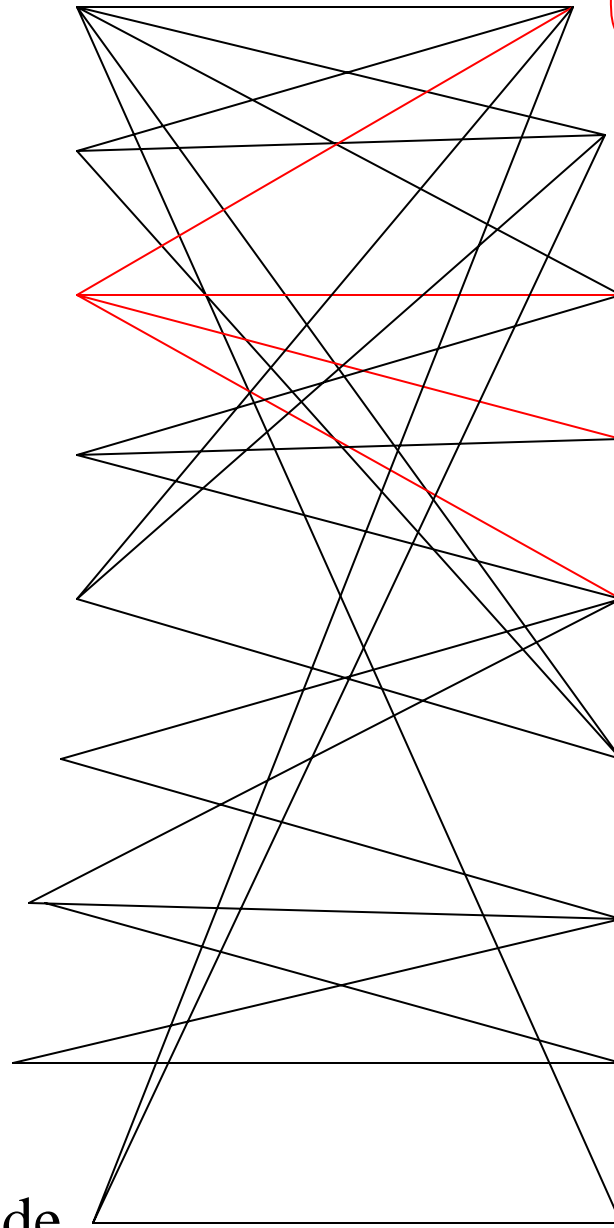
snill

yndig

pen

skjønn

velsmakende



good

nice

clever

sharp

cute

kind

pretty

beautiful

delicious

The
inverse
t-image
of
'cute'
I

god

hyggelig

flink

skarp

snill

yndig

pen

skjønn

velsmakende

good

nice

clever

sharp

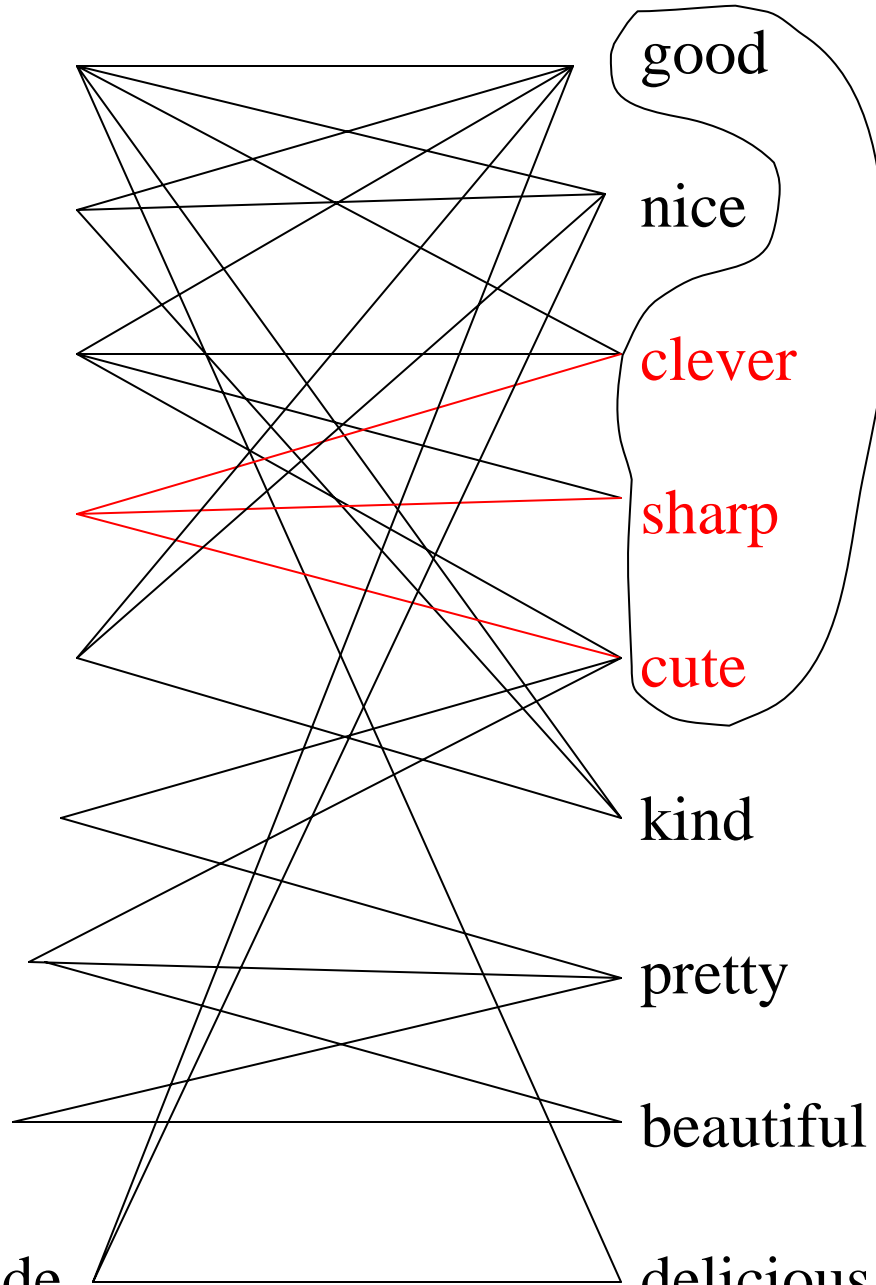
cute

kind

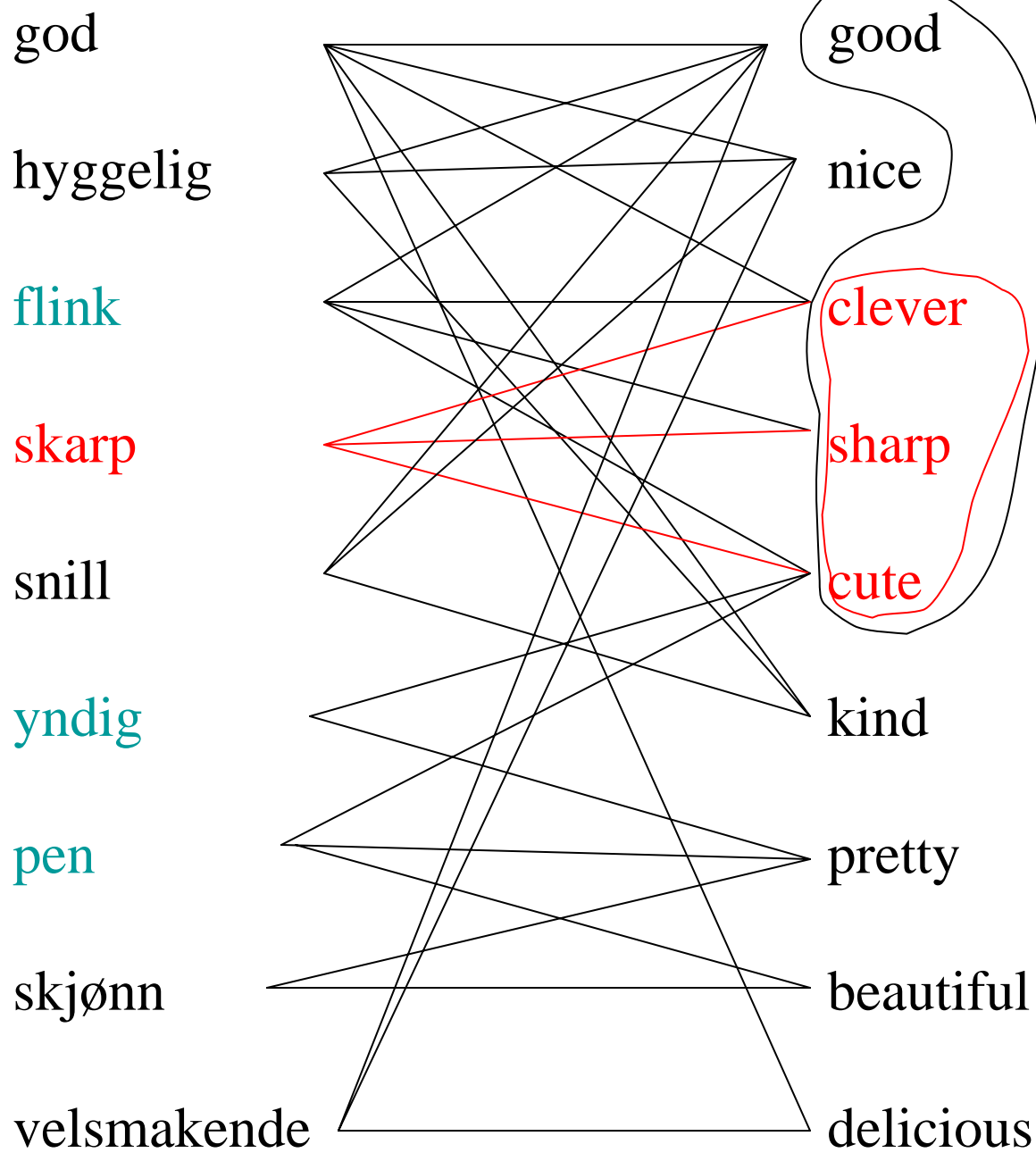
pretty

beautiful

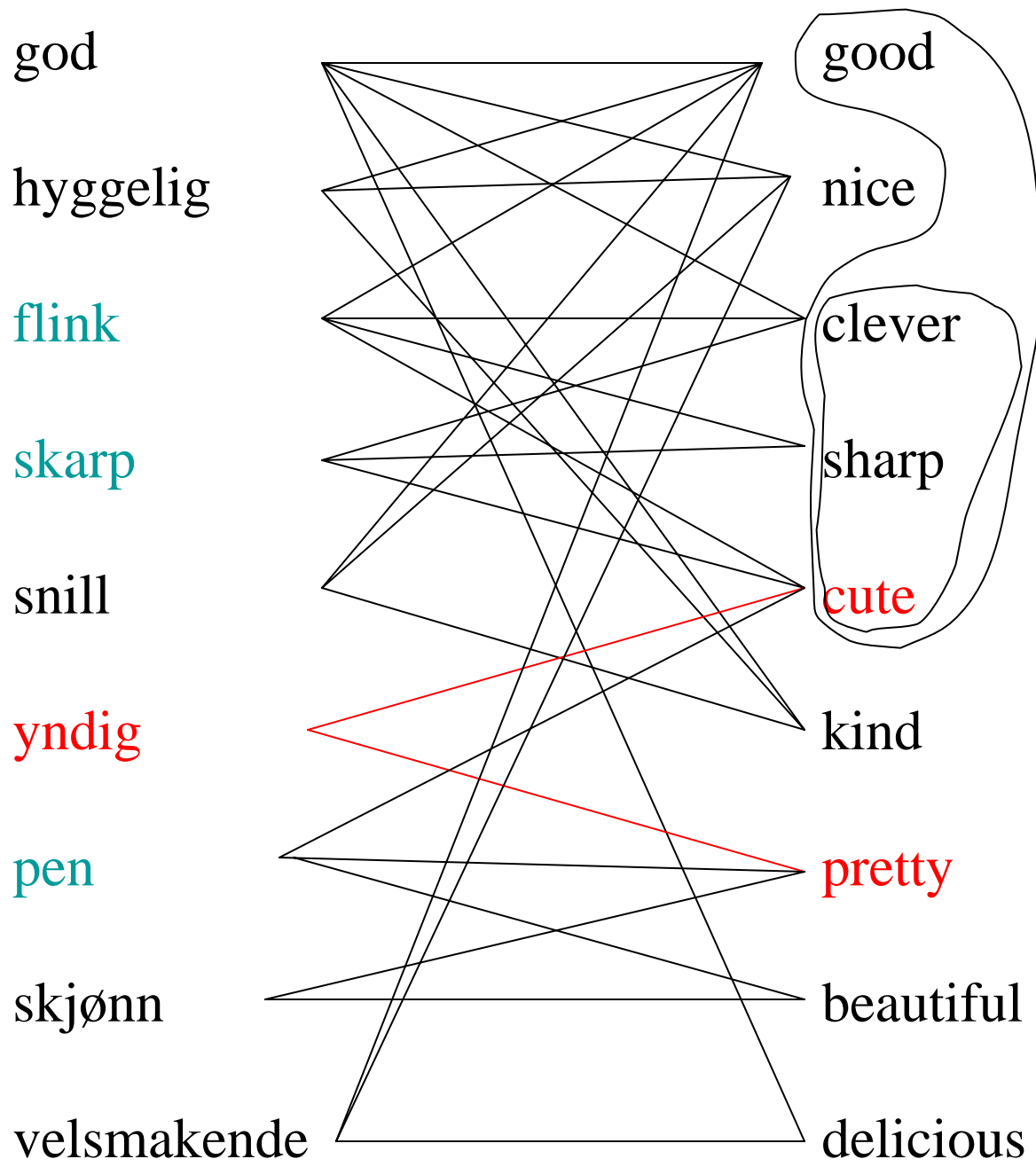
delicious



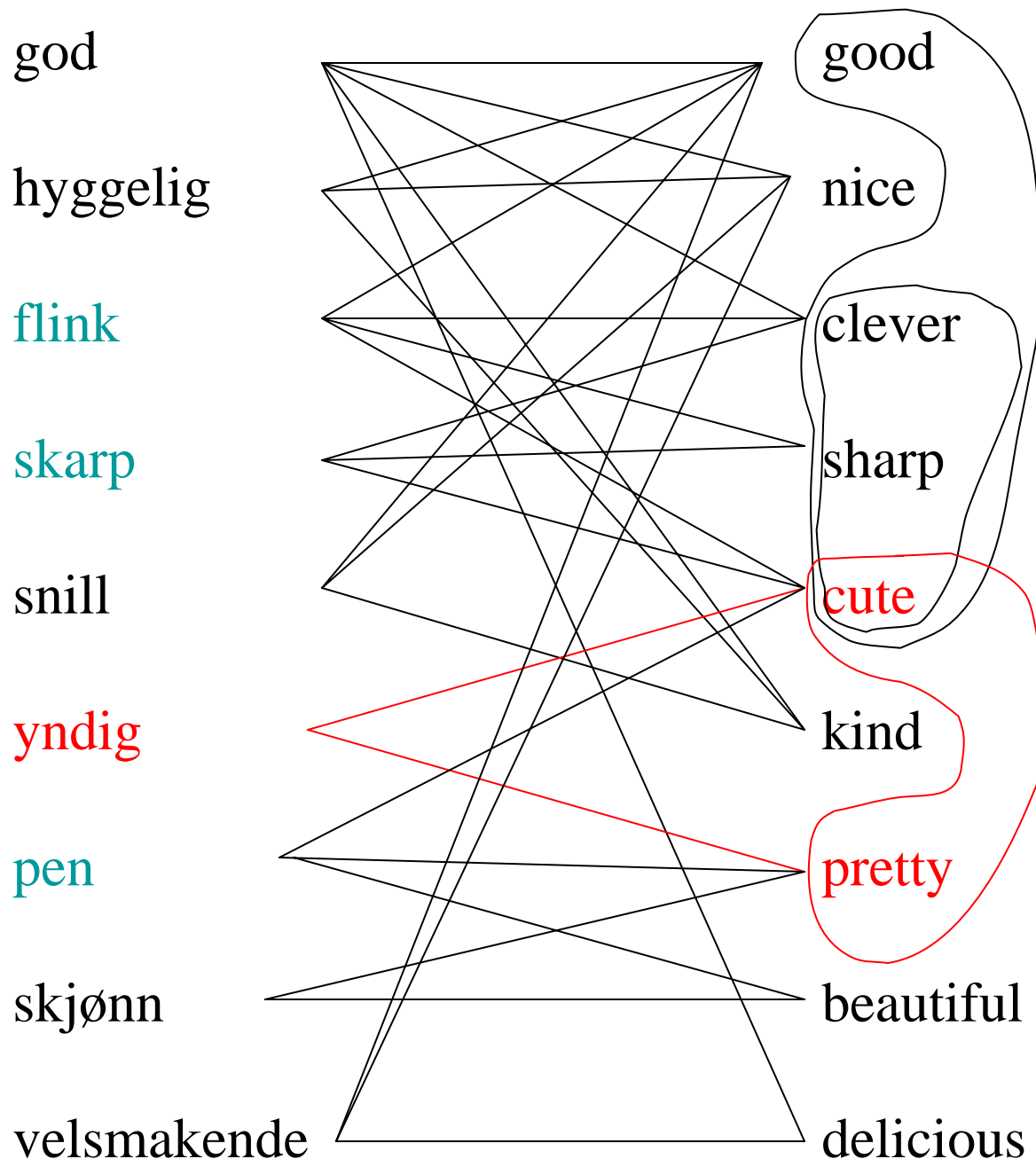
The
inverse
t-image
of
'cute'
II



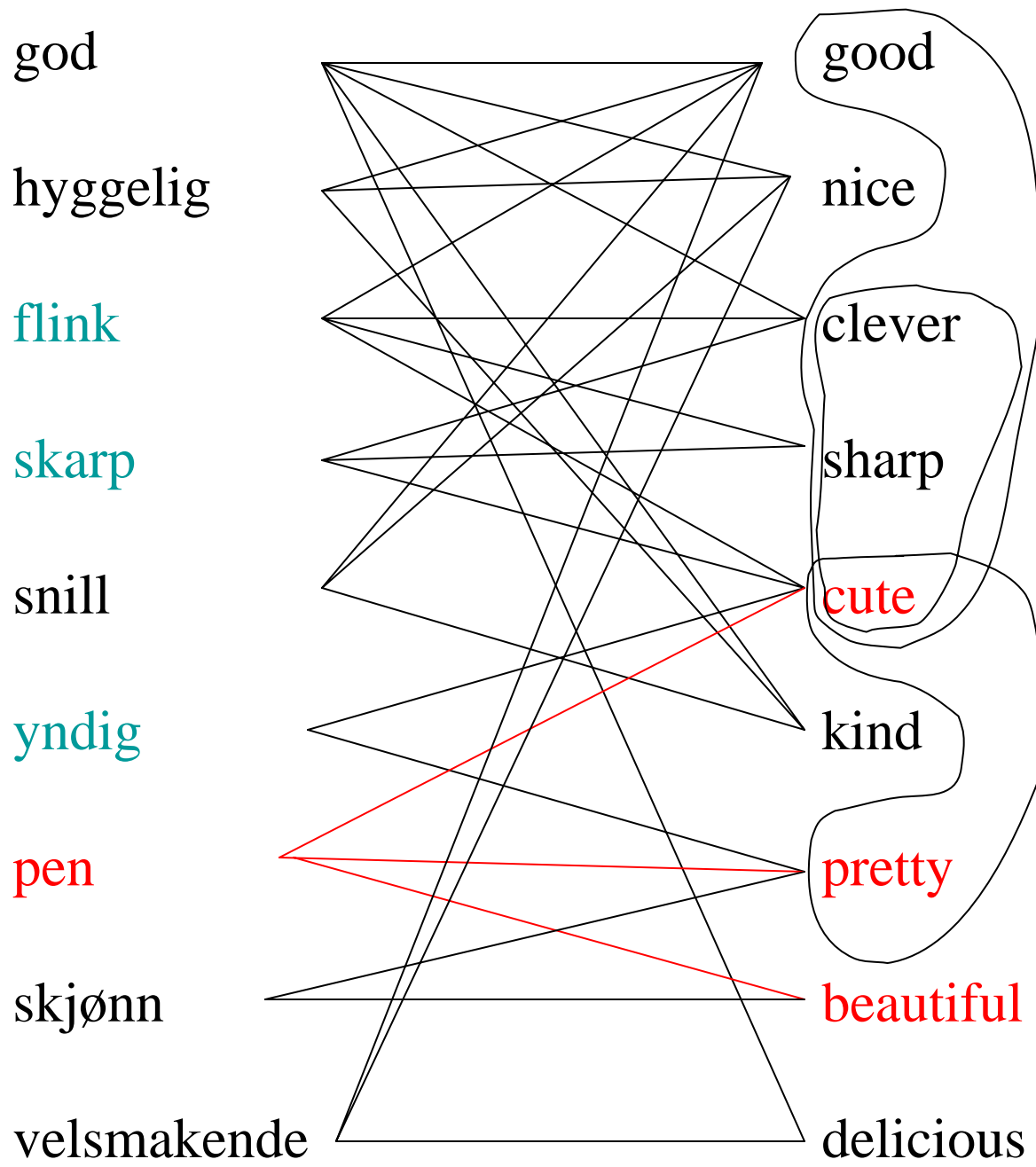
The
inverse
t-image
of
'cute'
II



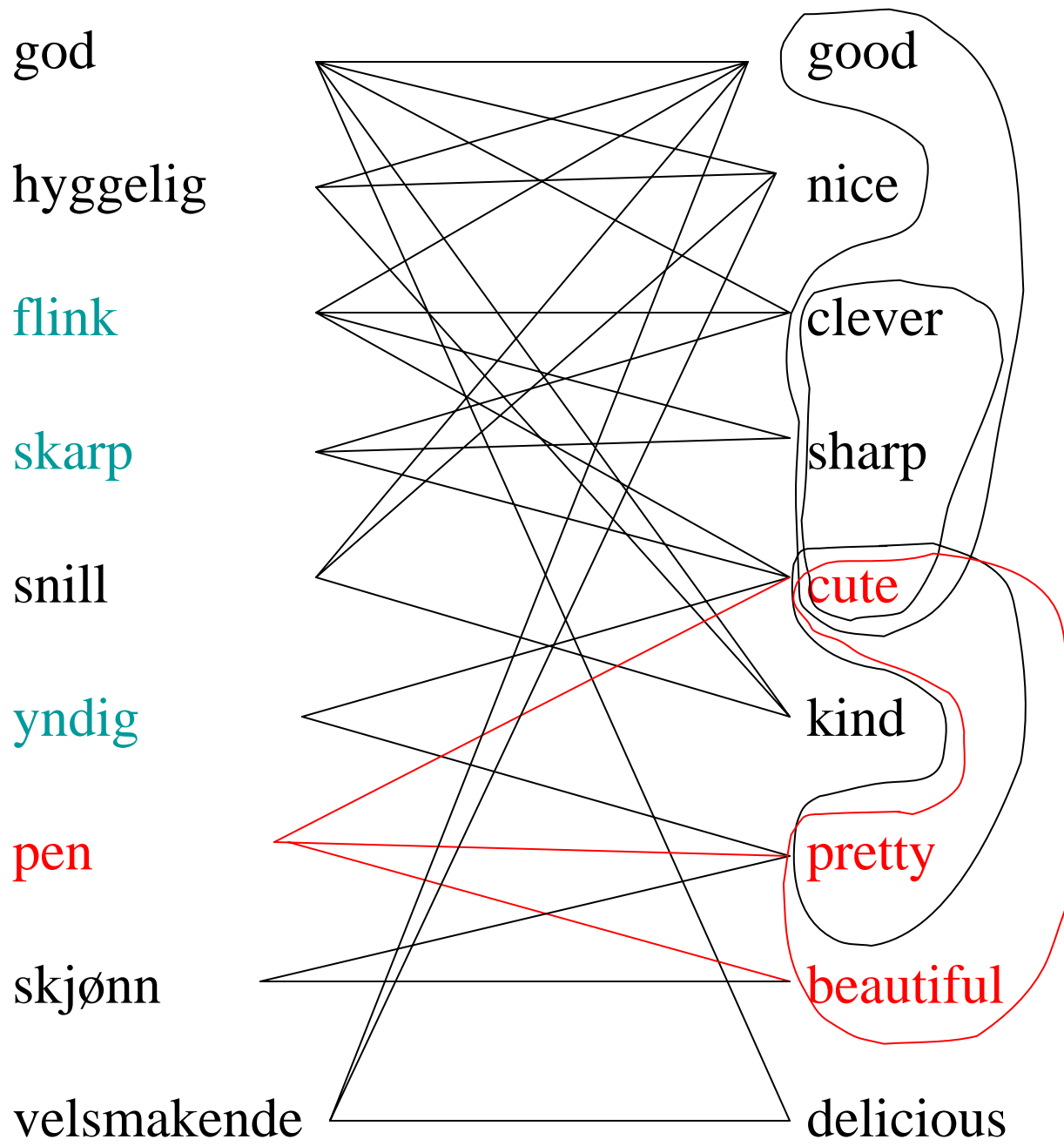
The
inverse
t-image
of
'cute'
III



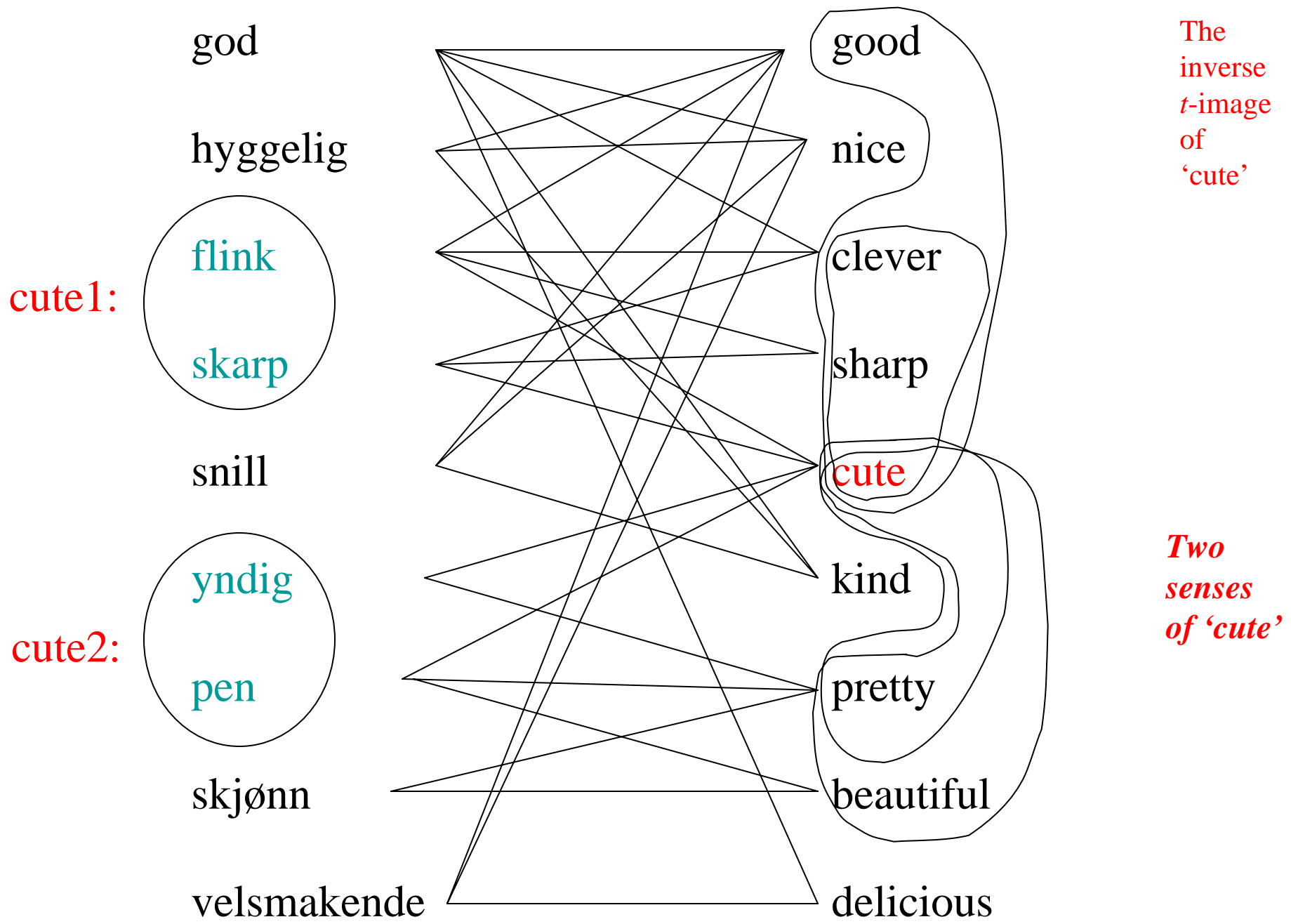
The
inverse
t-image
of
'cute'
III

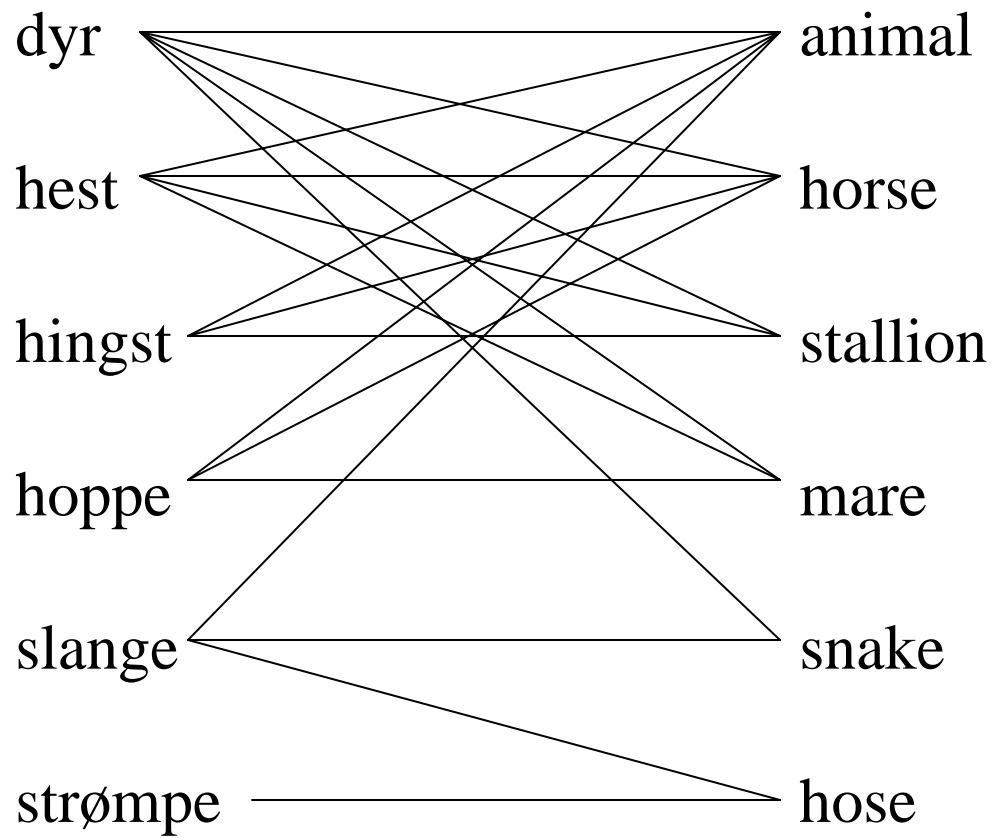


The
inverse
t-image
of
'cute'
IV



The
inverse
t-image
of
'cute'
IV





dyr

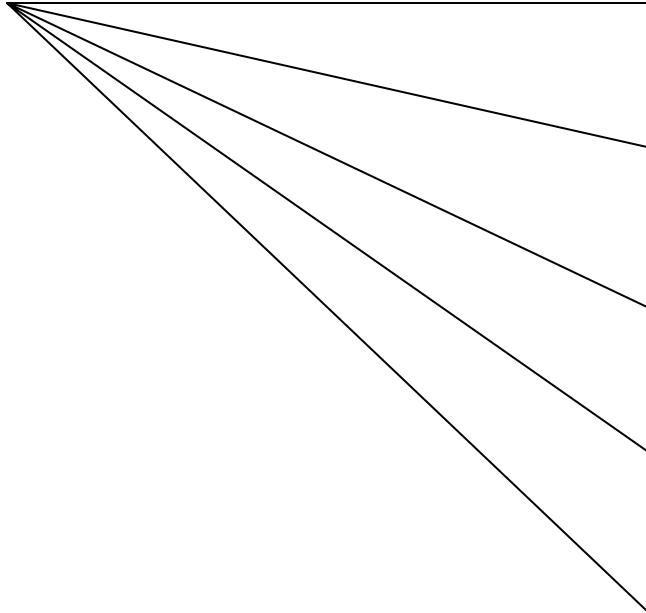
animal

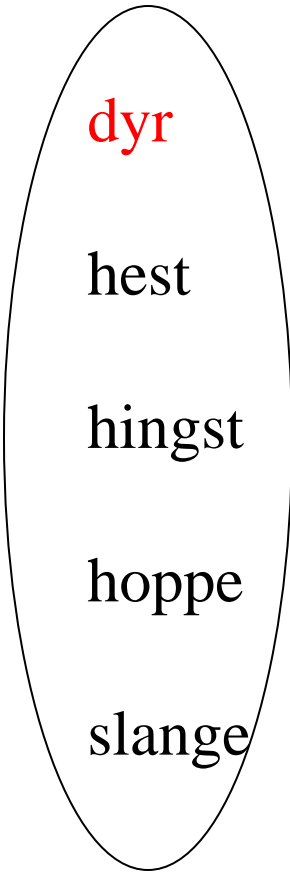
horse

stallion

mare

snake





dyr

hest

hingst

hoppe

slange

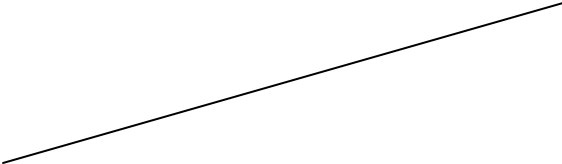
animal

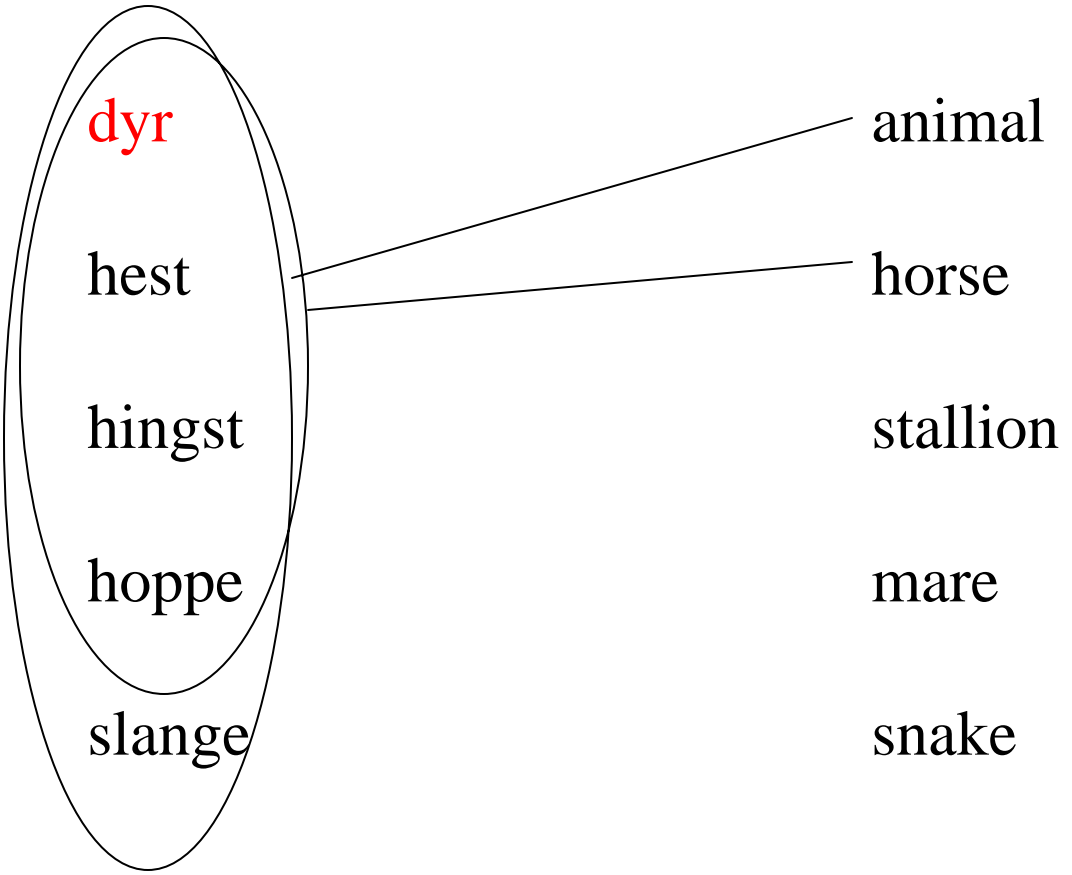
horse

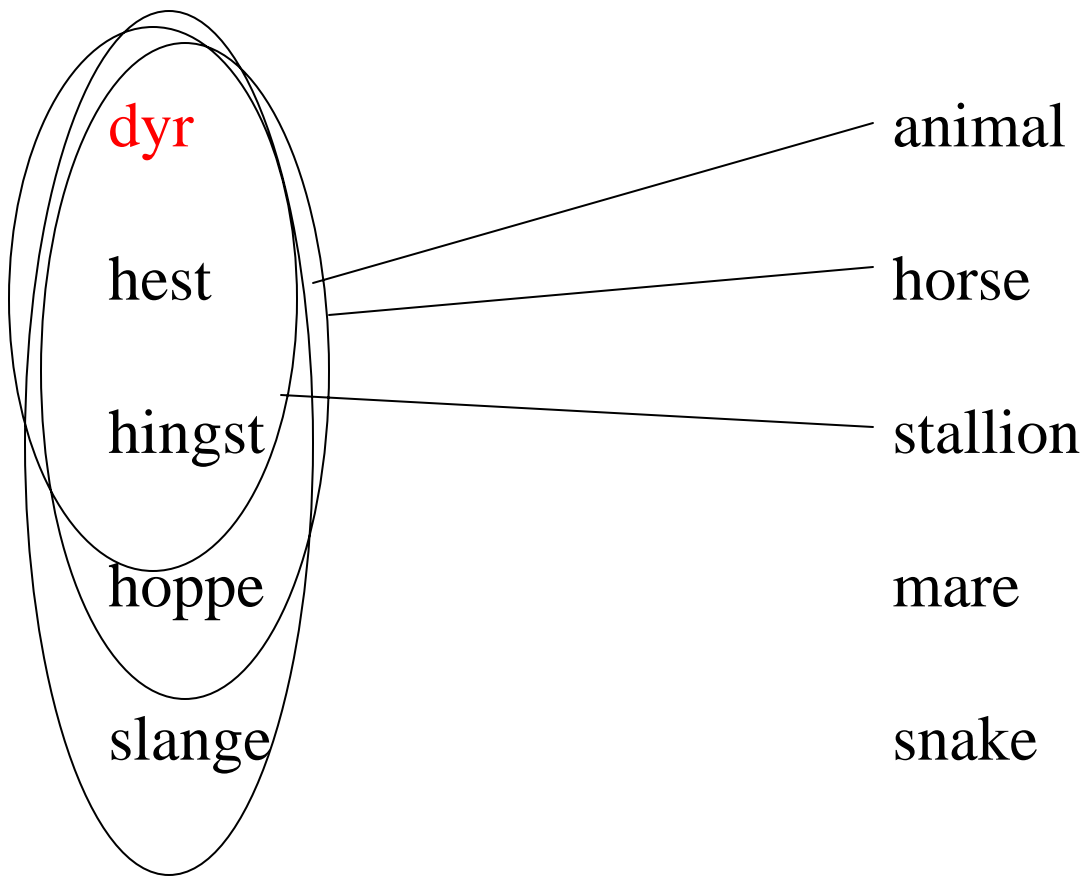
stallion

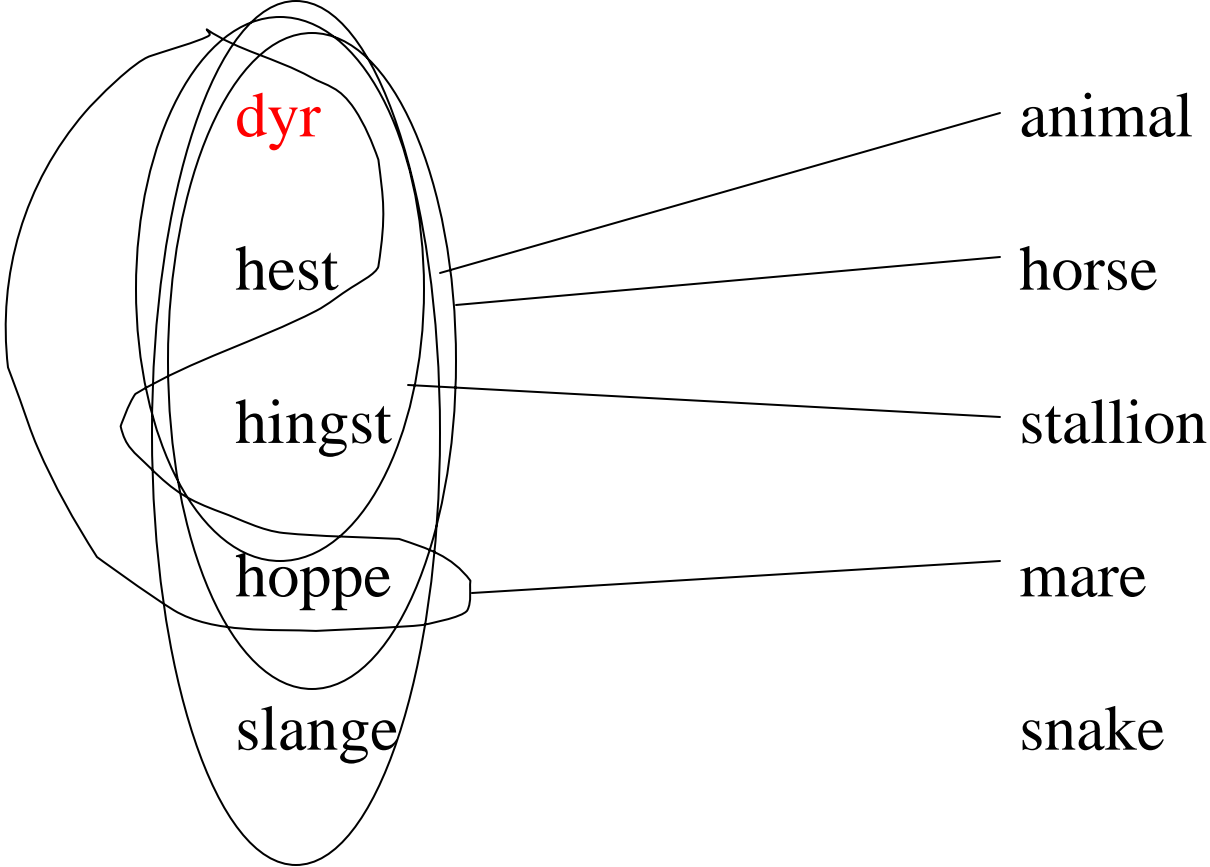
mare

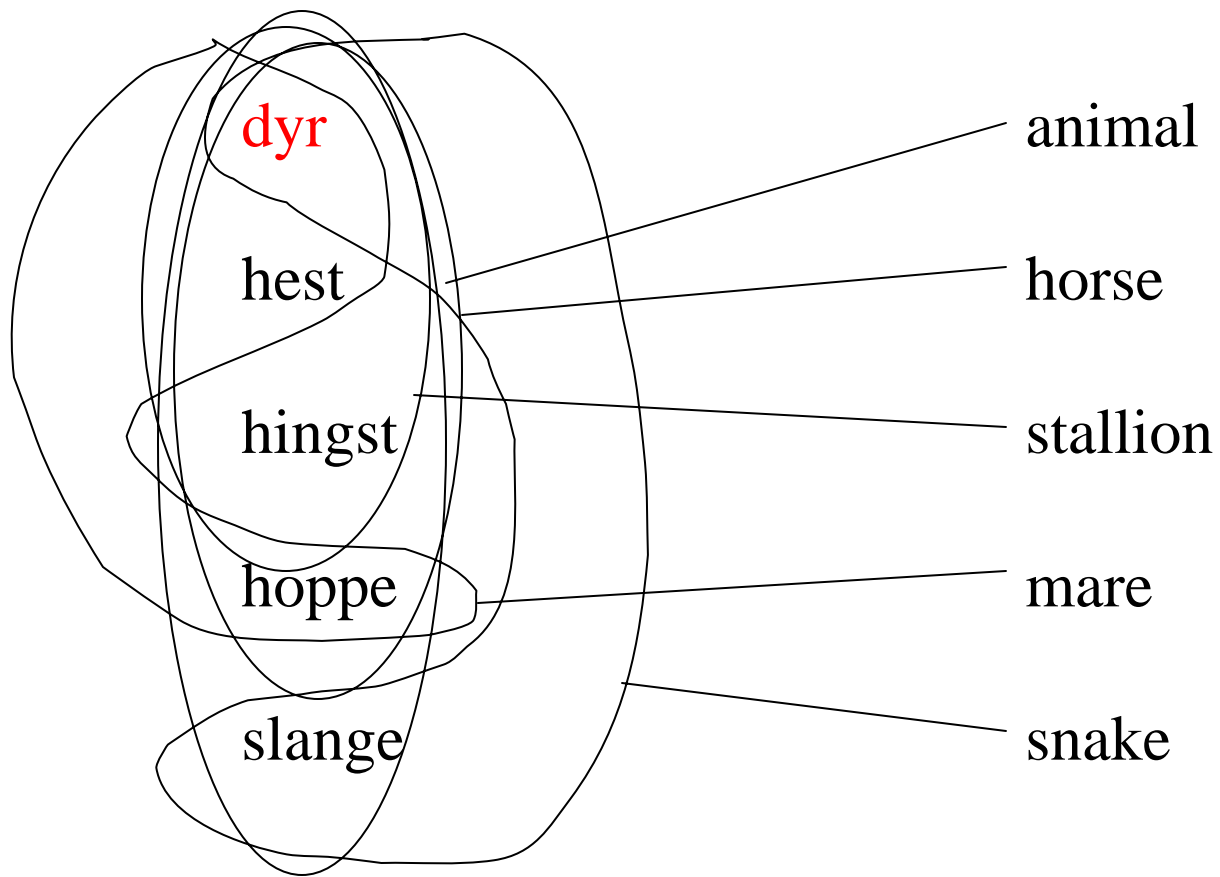
snake

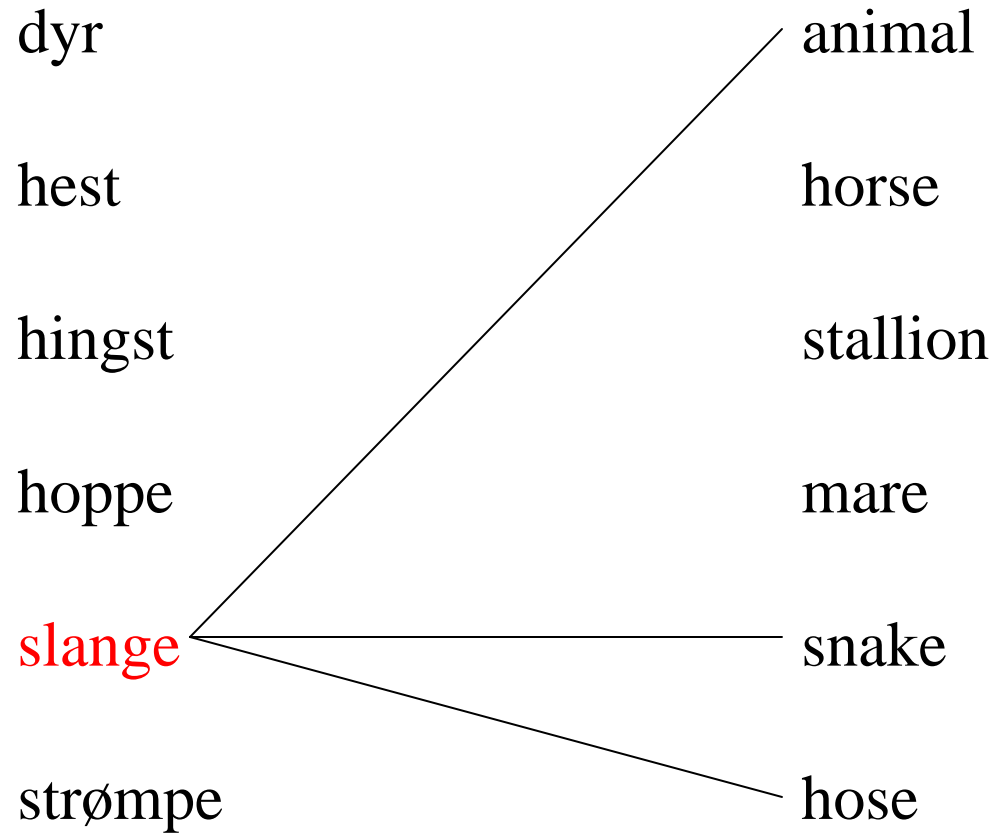


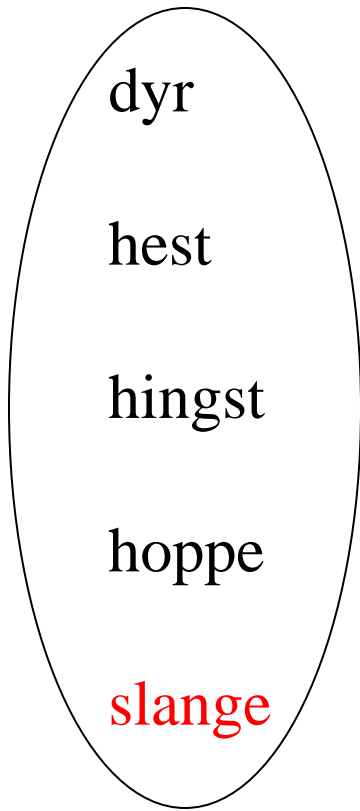








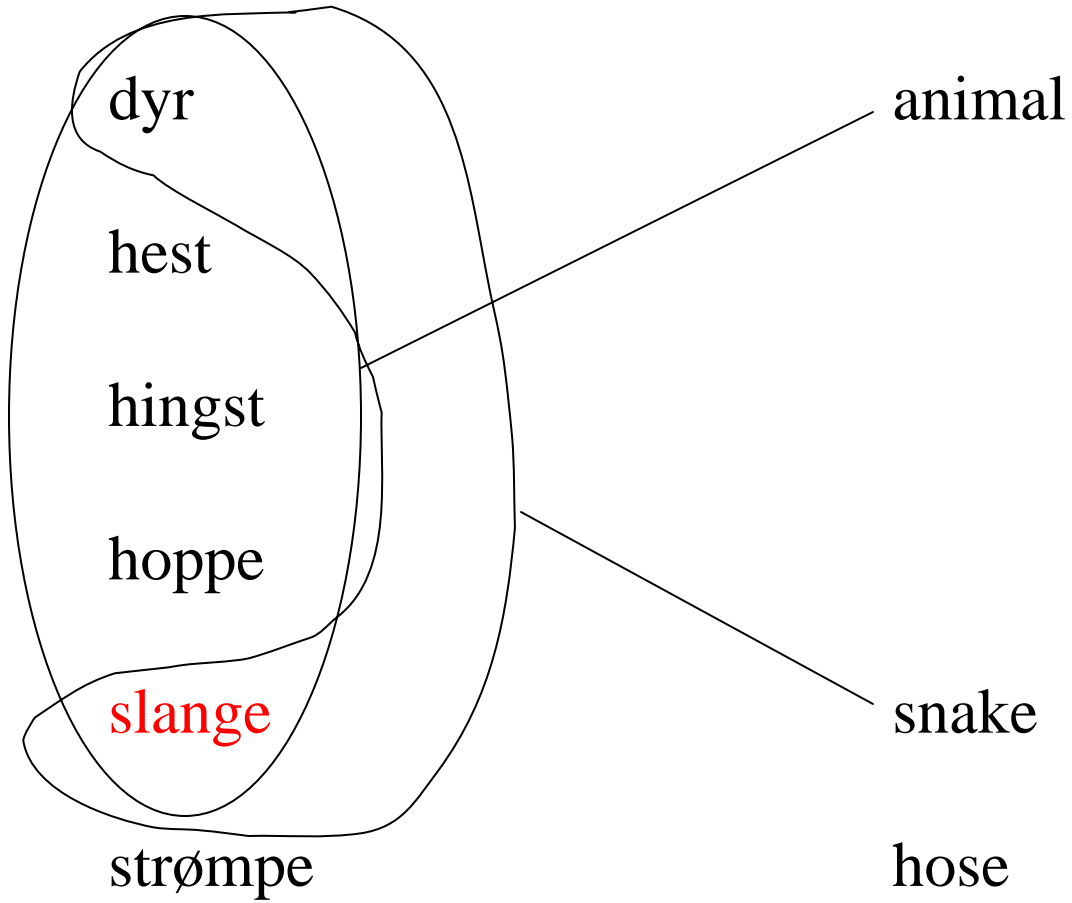


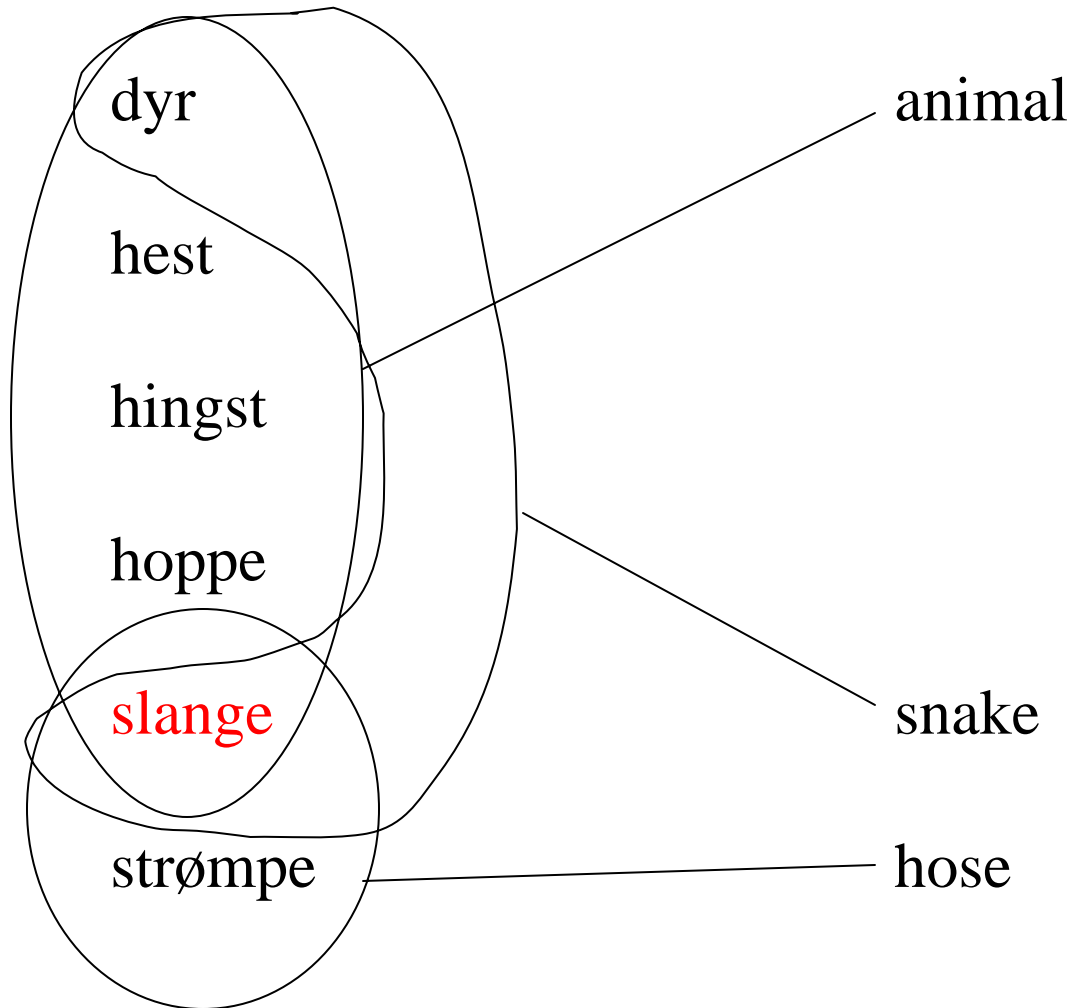


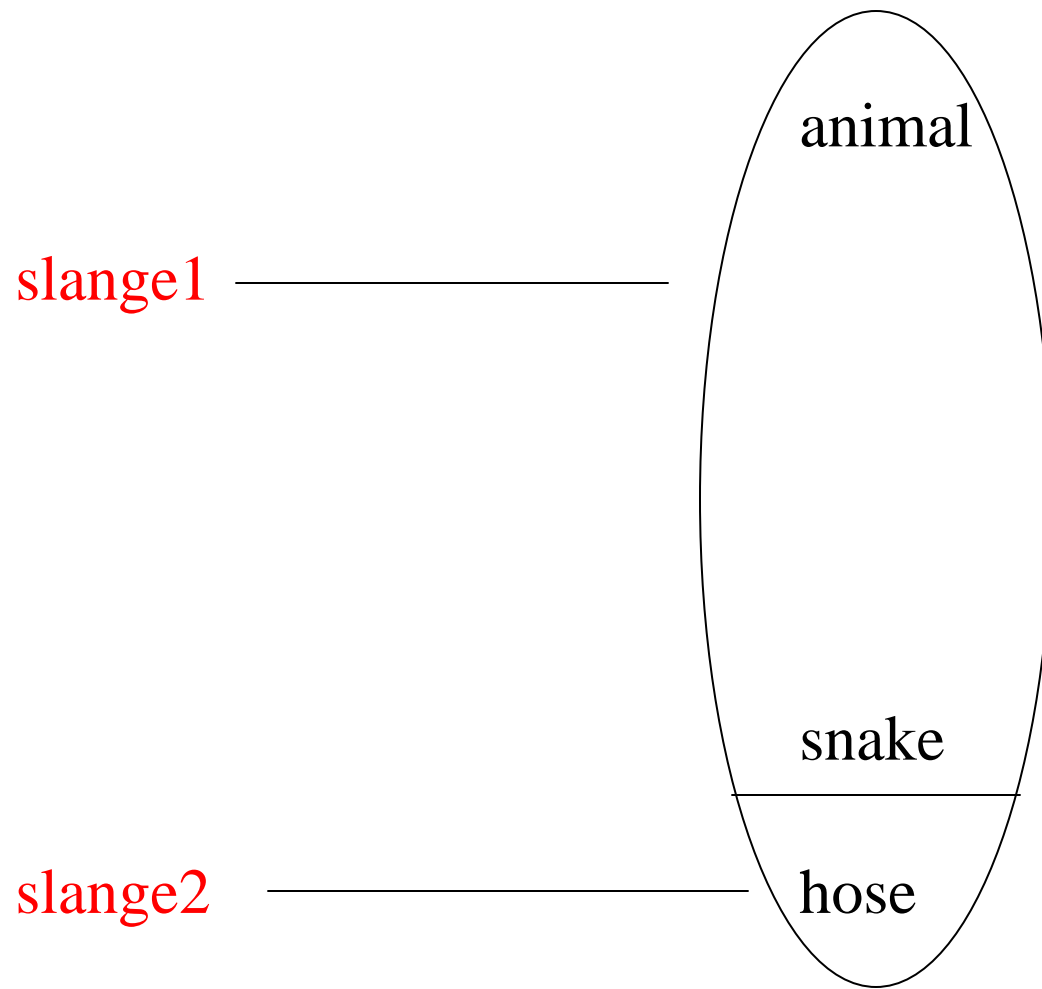
animal

snake

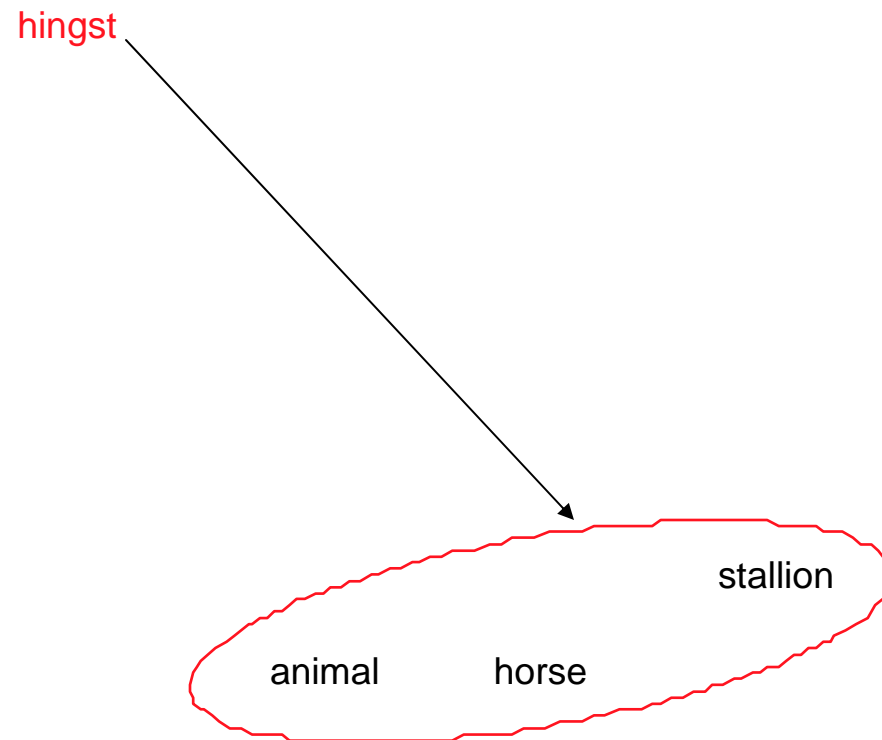
hose



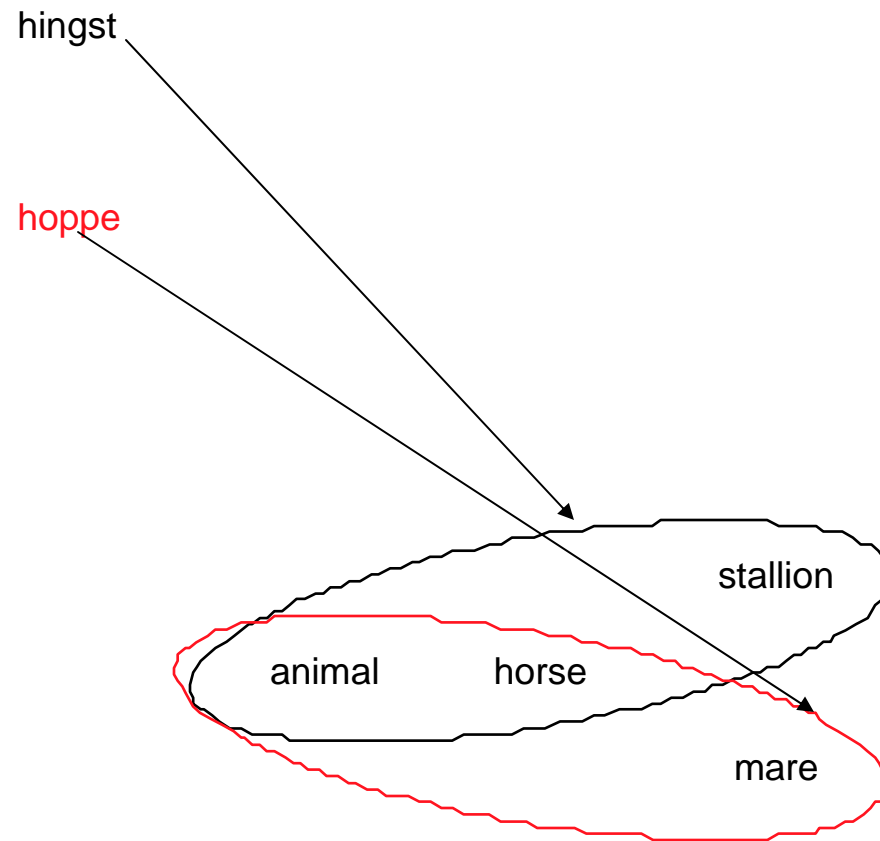




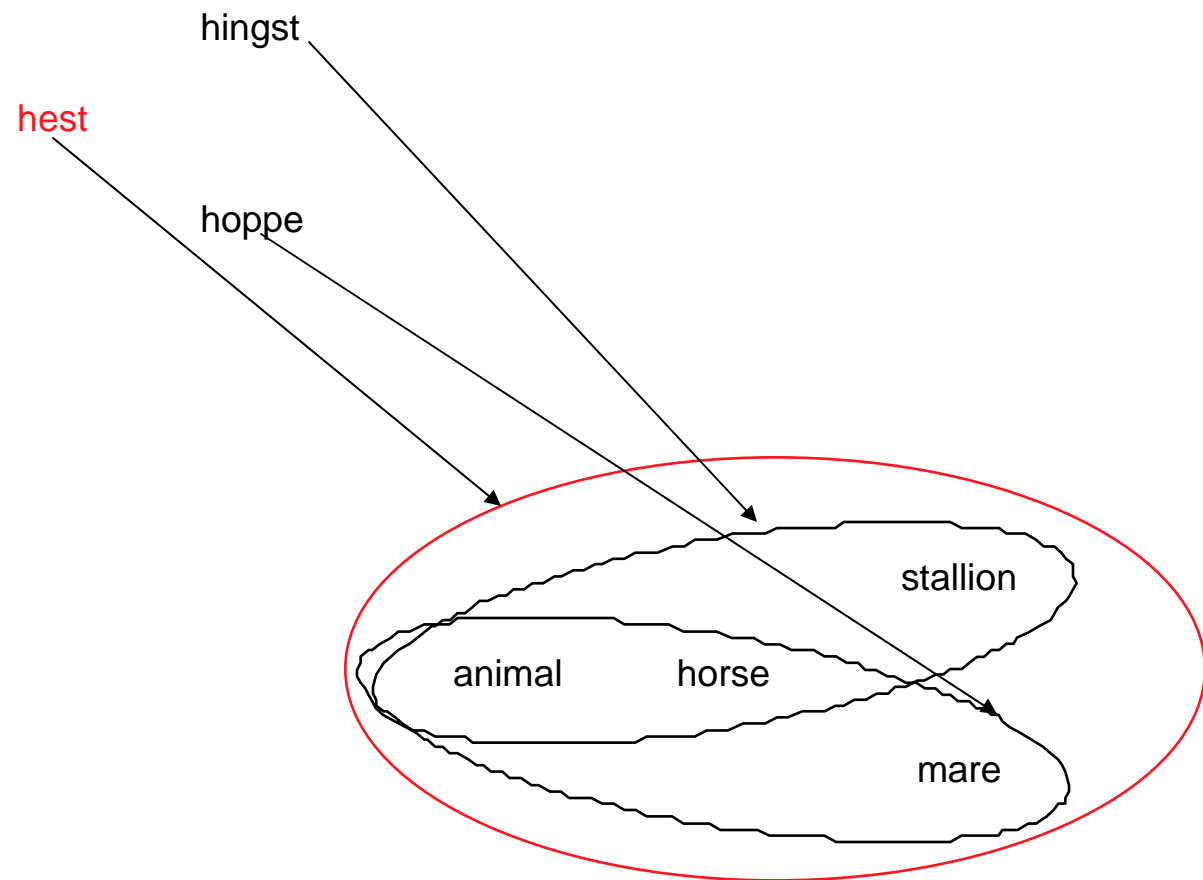
t-images within one semantic field:



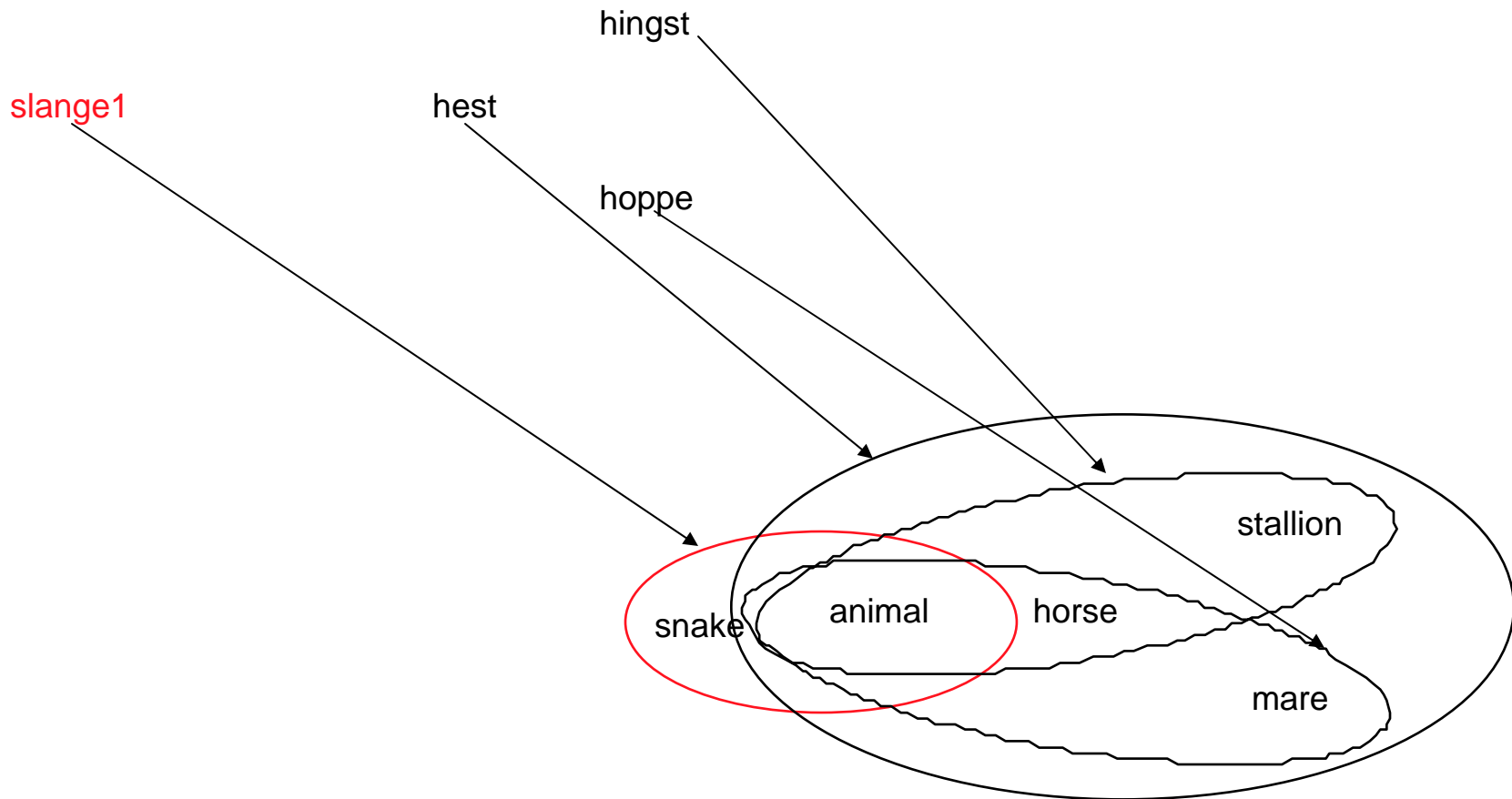
t-images within one semantic field:



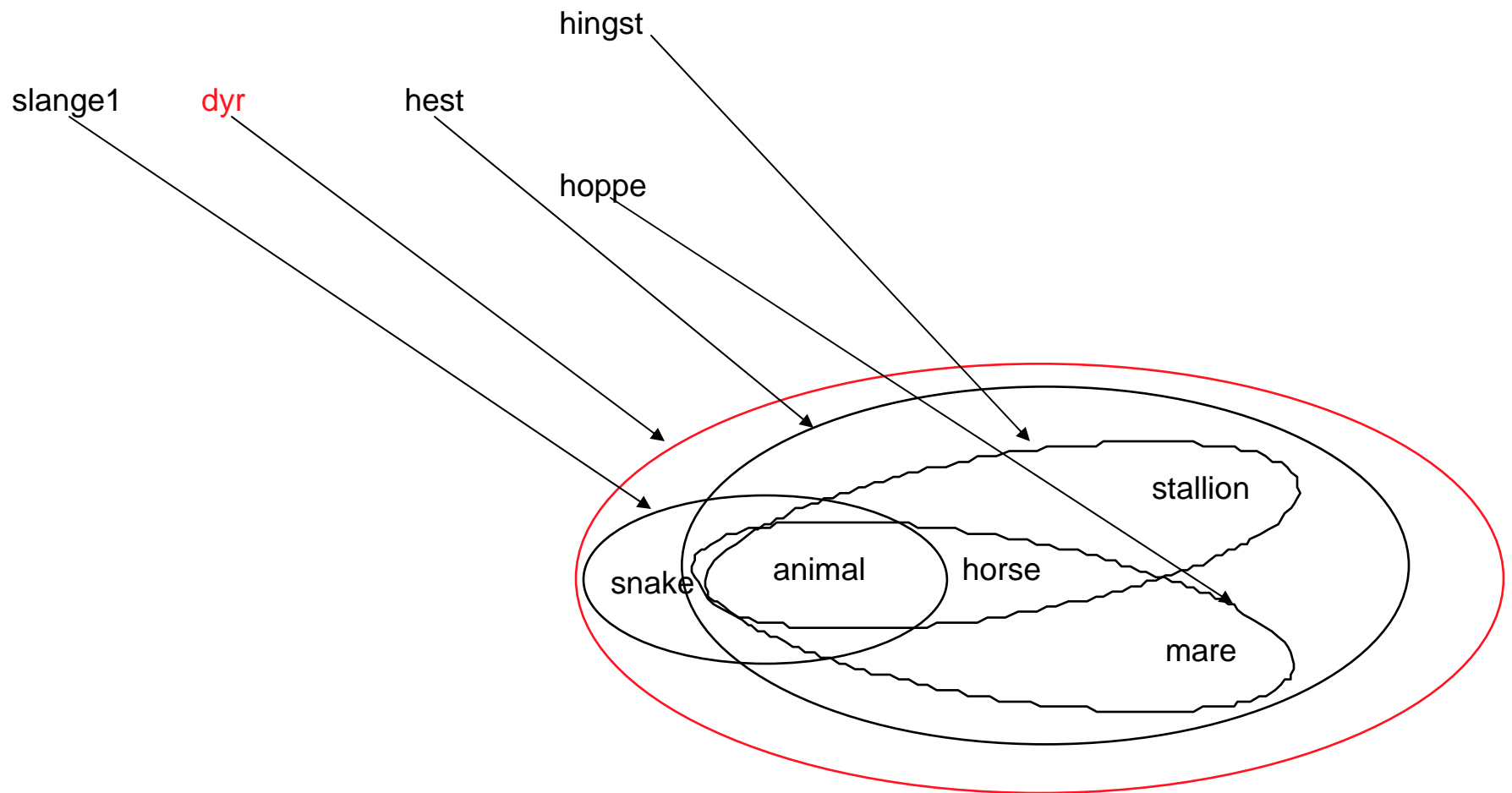
t-images within one semantic field:



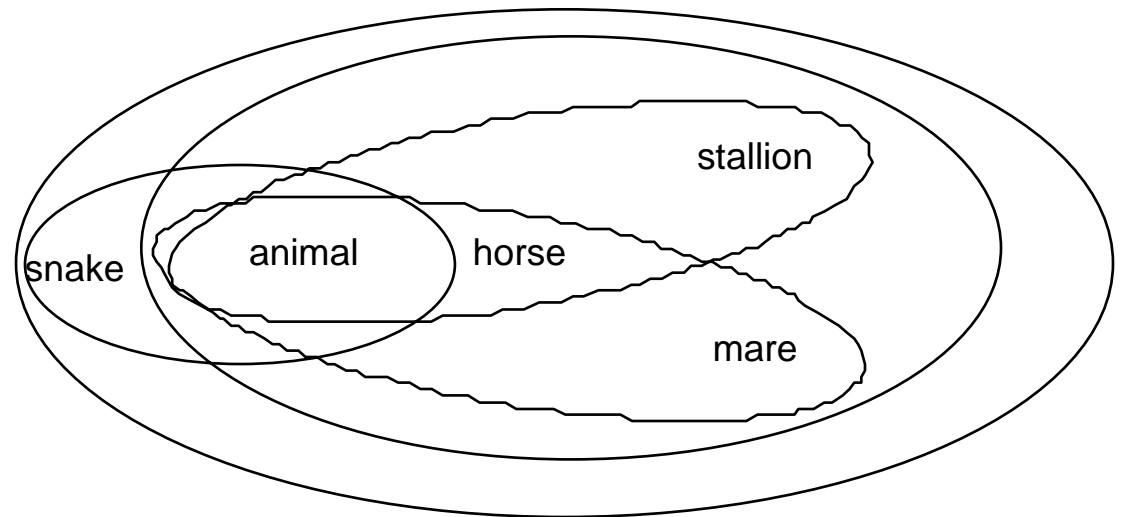
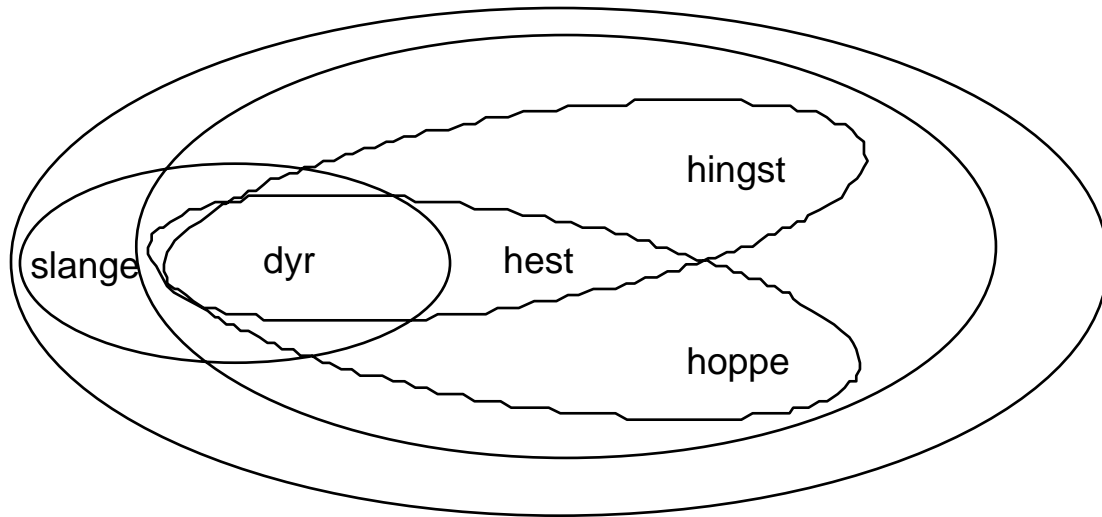
t-images within one semantic field:



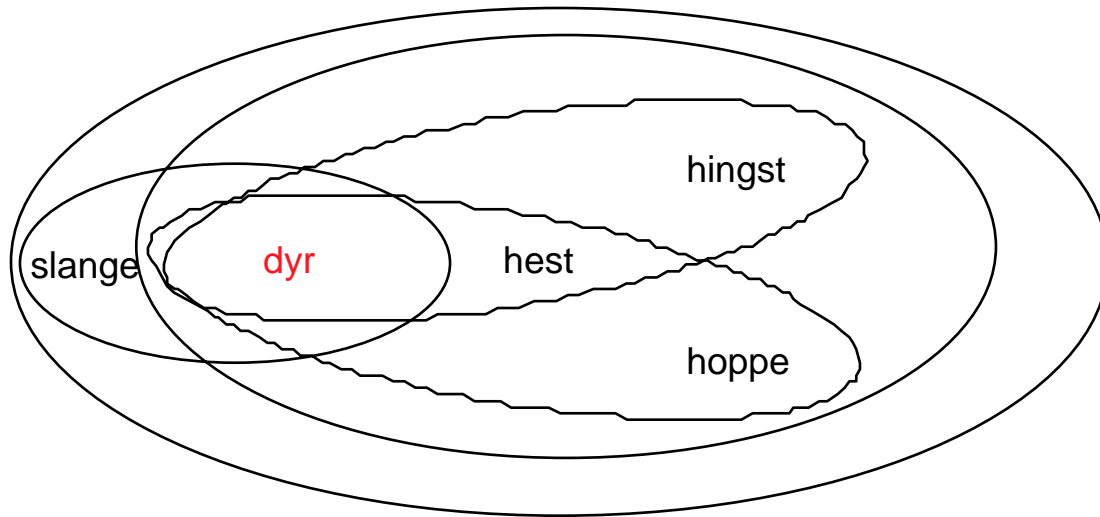
t-images within one semantic field:



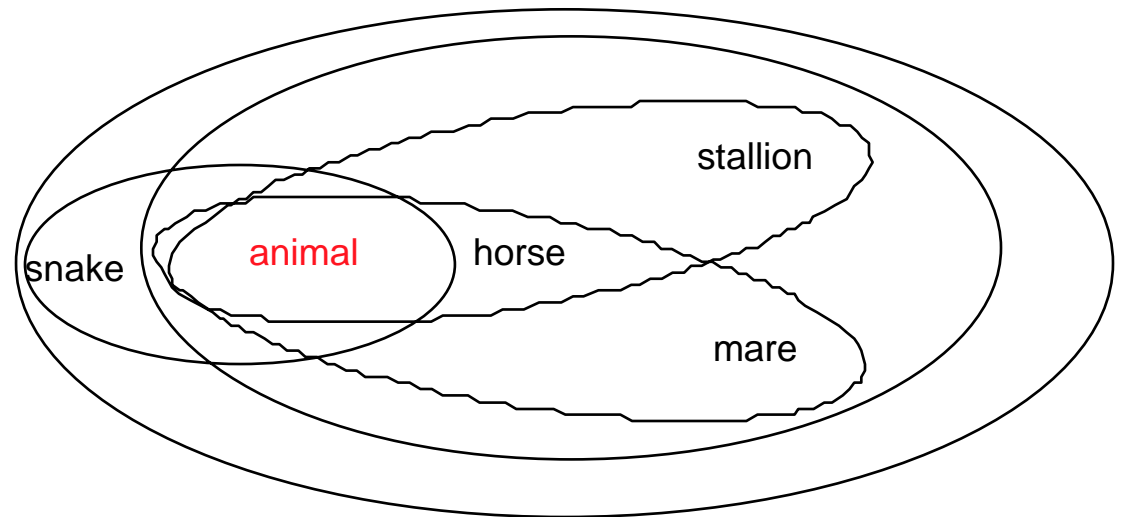
Corresponding semantic fields assign set structures to each other:



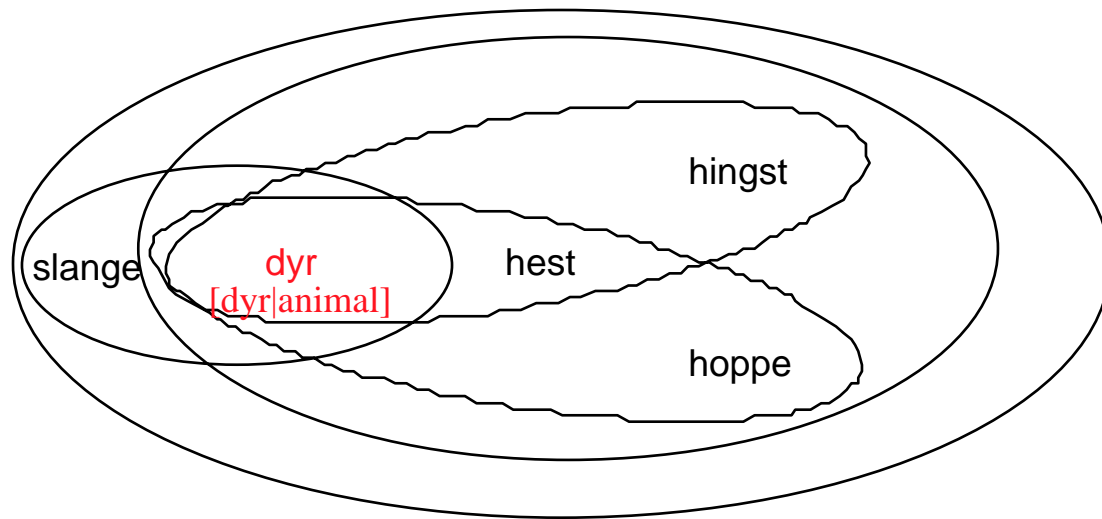
Semantic features are constructed from the peaks:



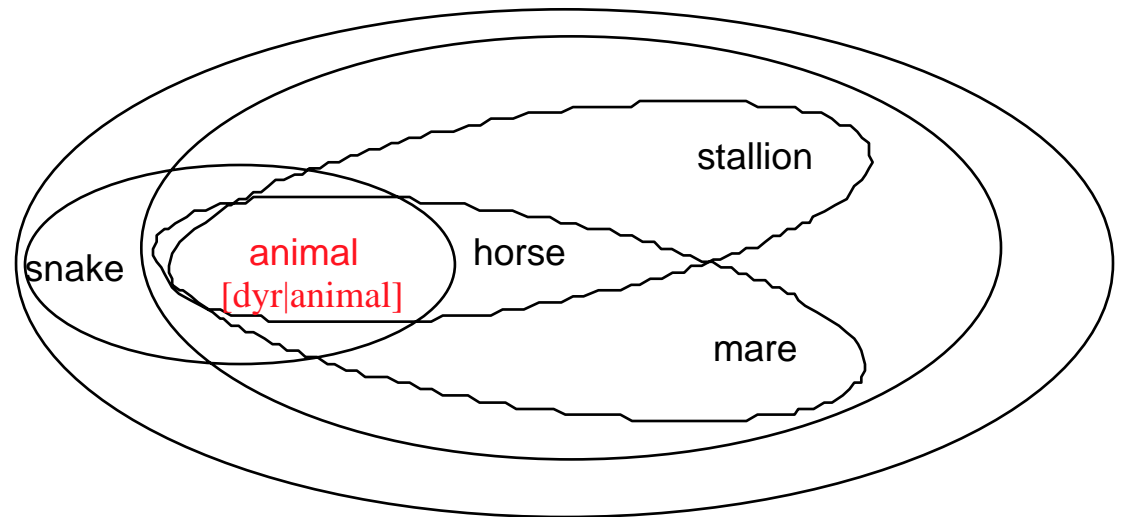
[dyr|animal]



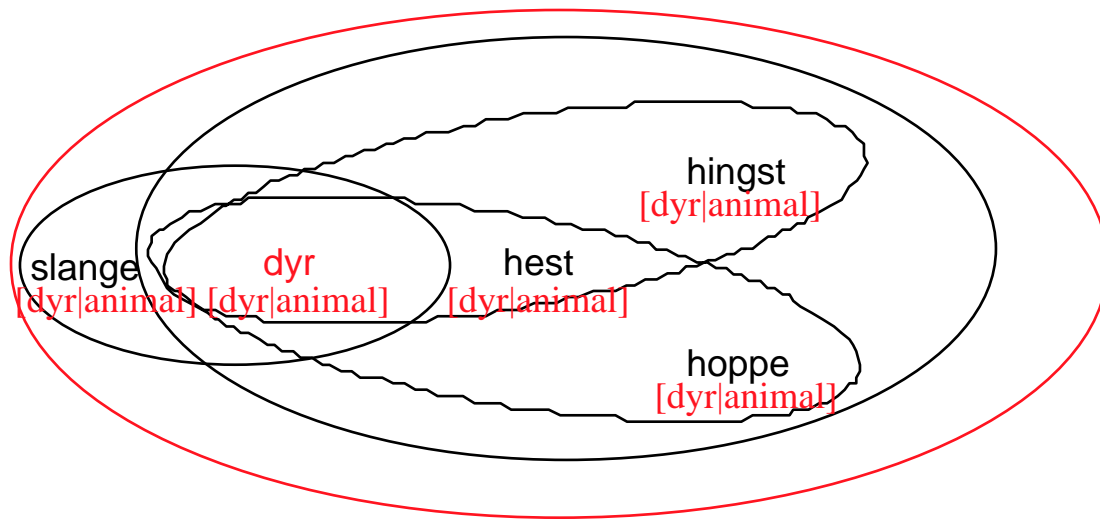
The feature is assigned to the two senses from which it is constructed...



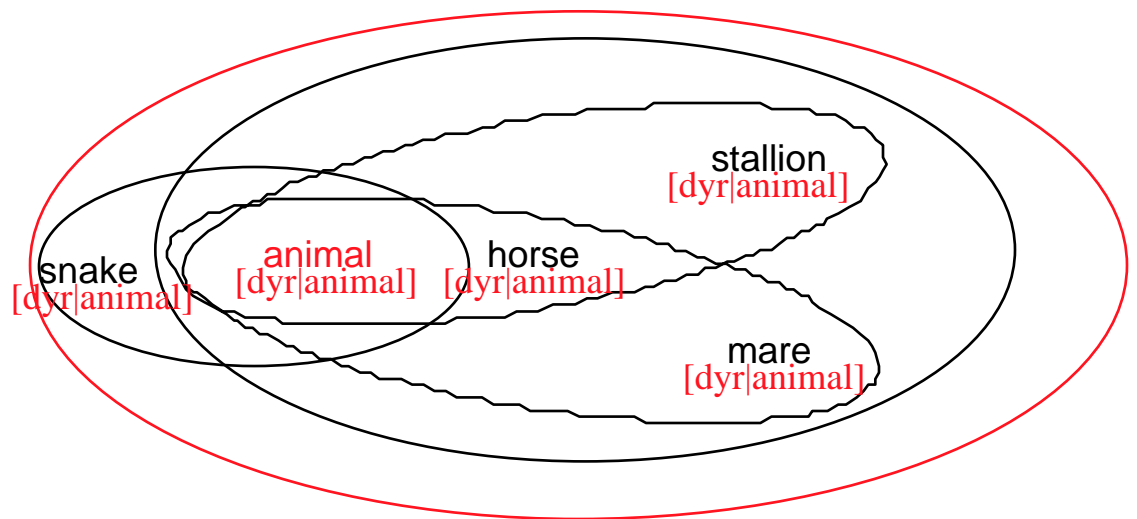
[dyr|animal]



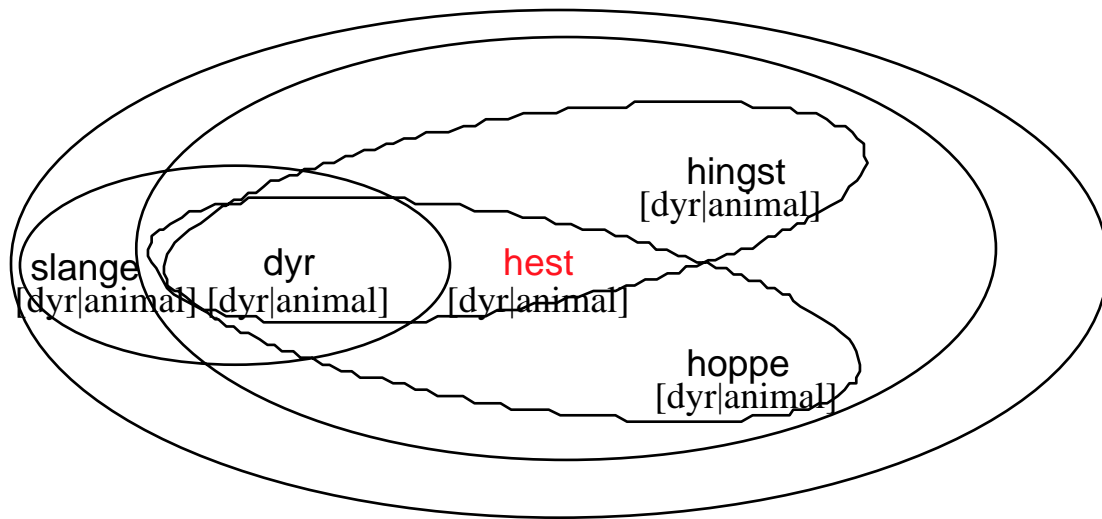
...and besides to all the lower senses in the *t*-images of 'dyr'
and 'animal':



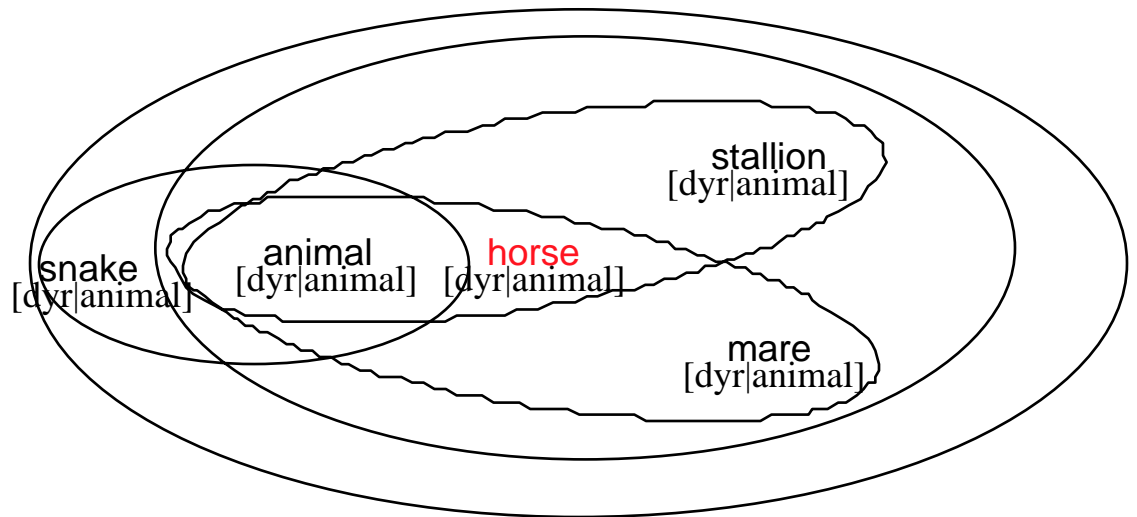
[dyr|animal]



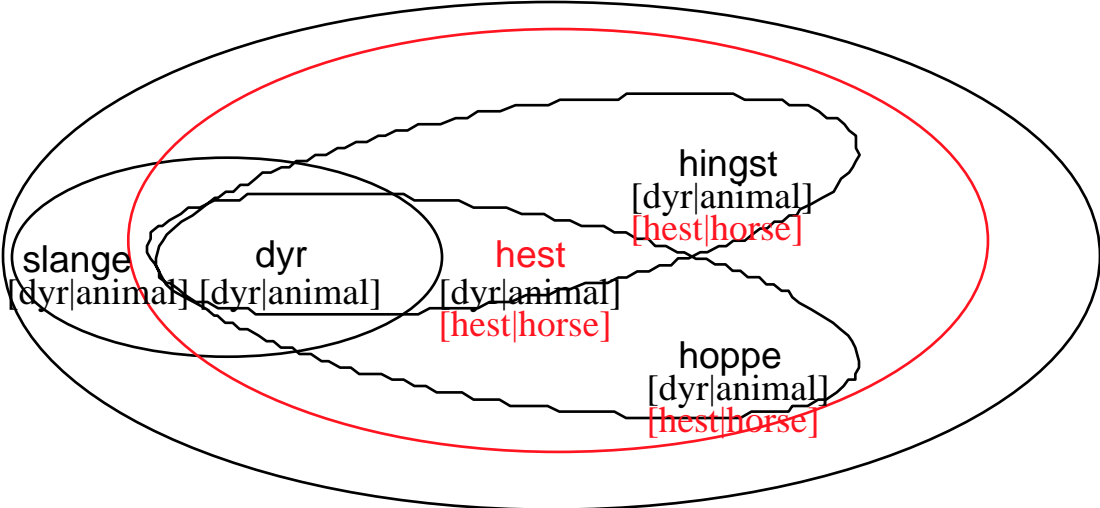
We go on to the lower peaks:



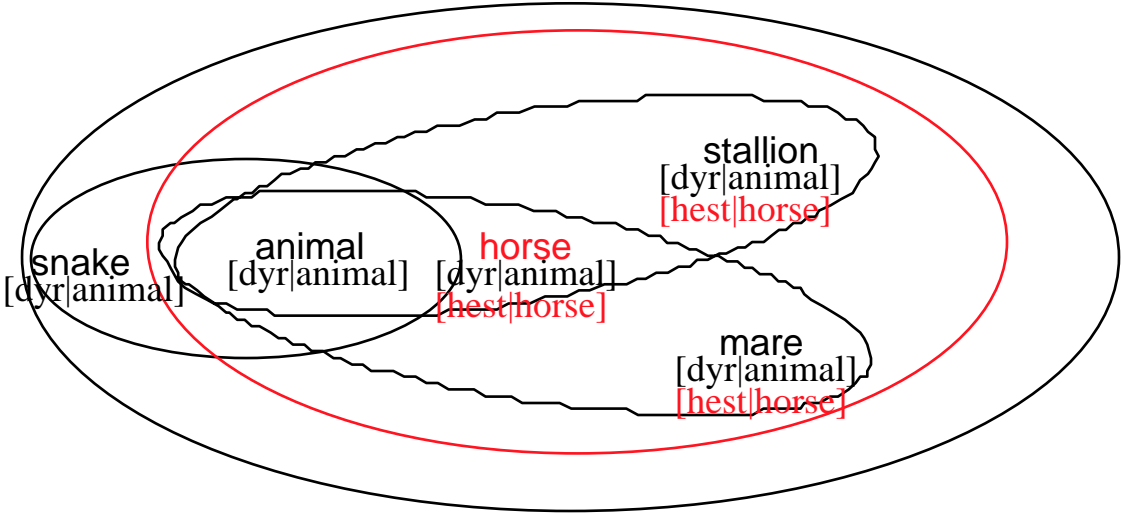
[dyr|animal]
[hest|horse]



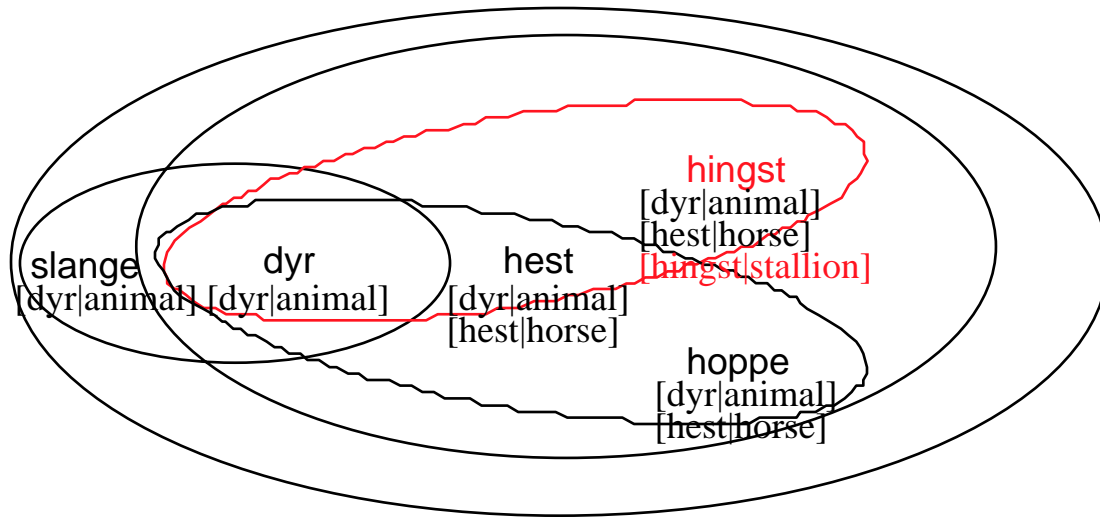
The feature is assigned to 'hest', 'horse' and lower senses in the *t*-images of 'hest' and 'horse':



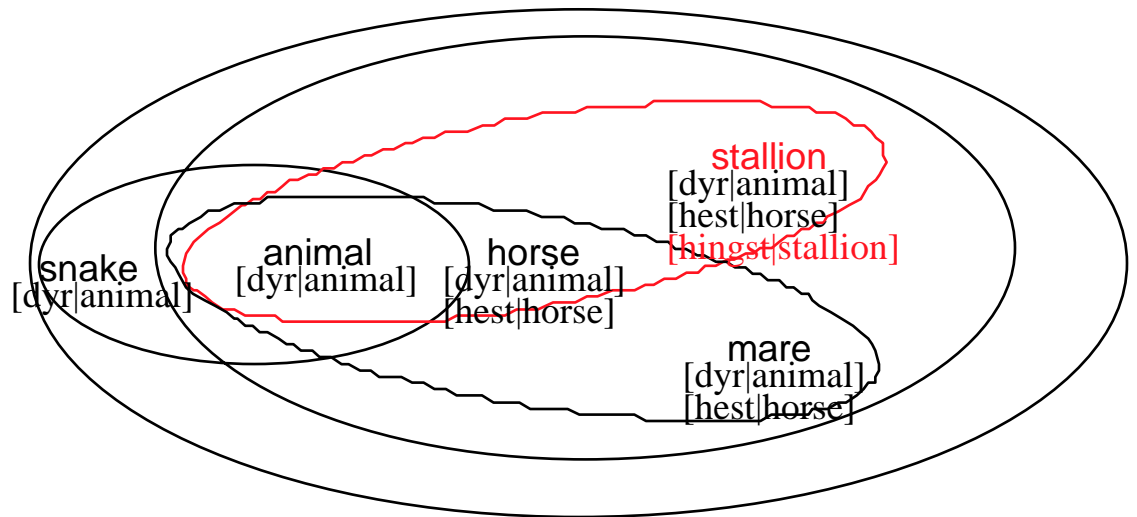
[dyr|animal]
[hest|horse]



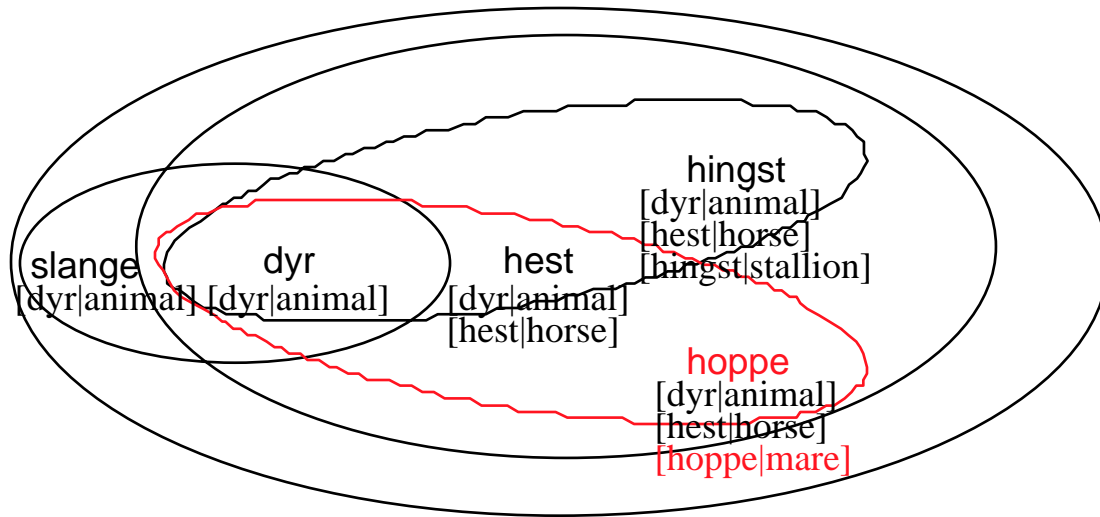
The two next peaks have the same height –
we arbitrarily choose ‘hingst’/‘stallion’:



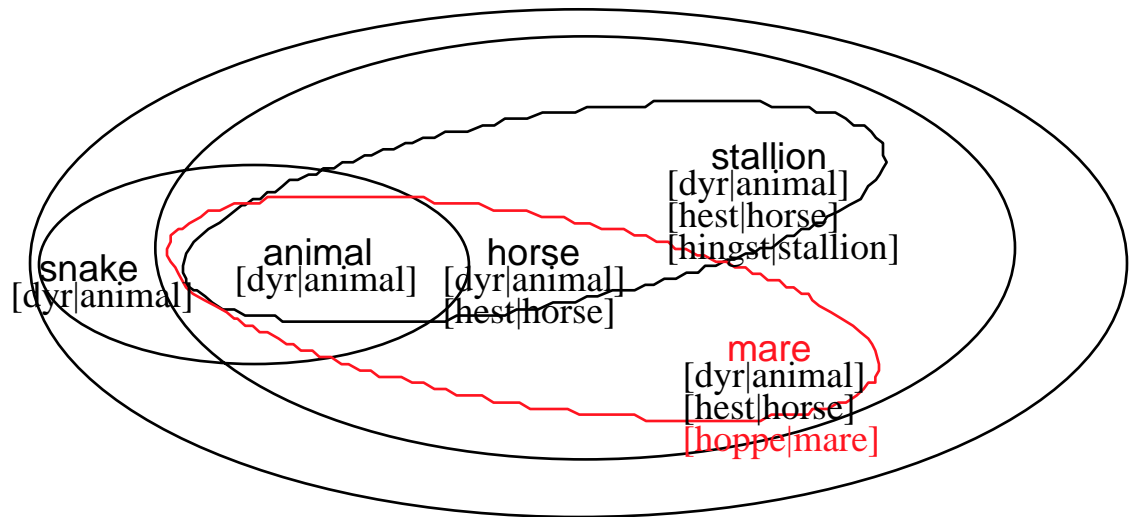
[dyr|animal]
[hest|horse]
[hingst|stallion]



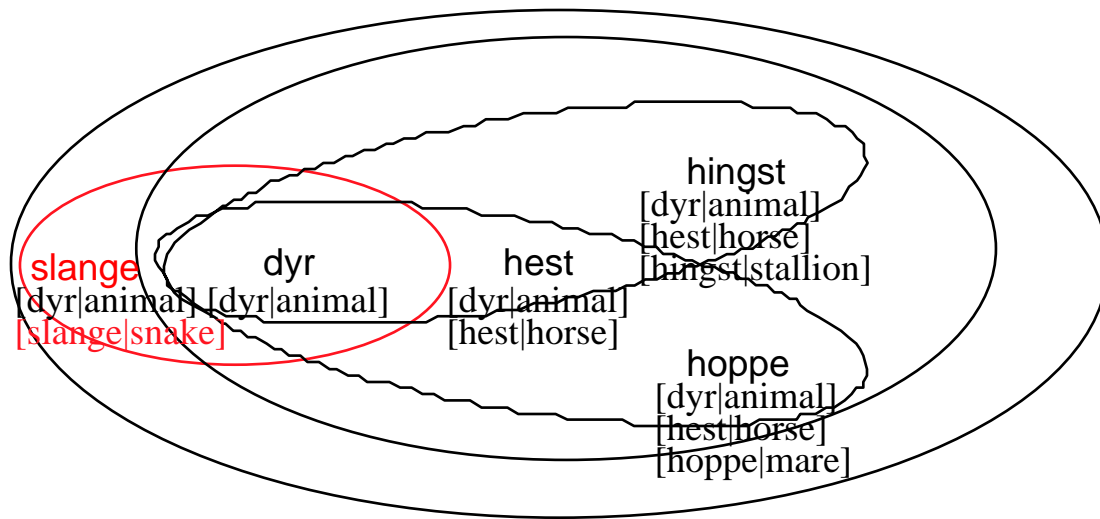
Then 'hoppe'/'mare':



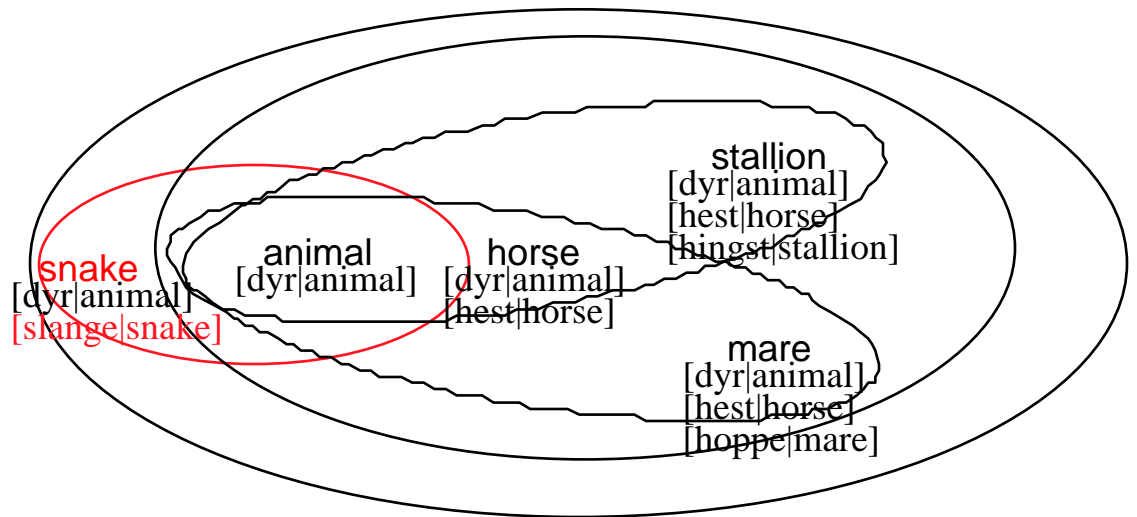
[dyr|animal]
[hest|horse]
[hingst|stallion]
[hoppe|mare]



And finally 'slange'/'snake':

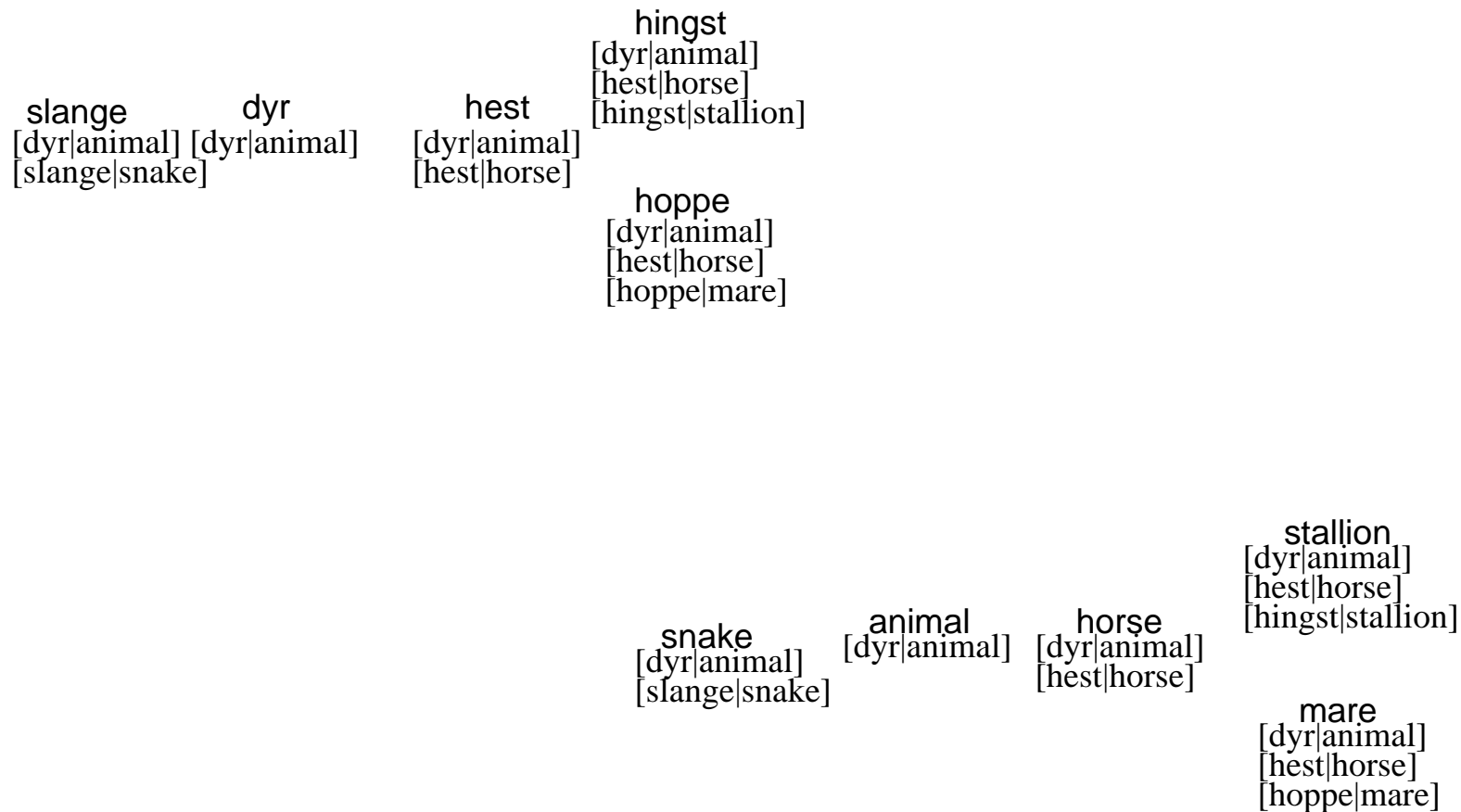


[dyr|animal]
[hest|horse]
[hingst|stallion]
[hoppe|mare]
[slange|snake]

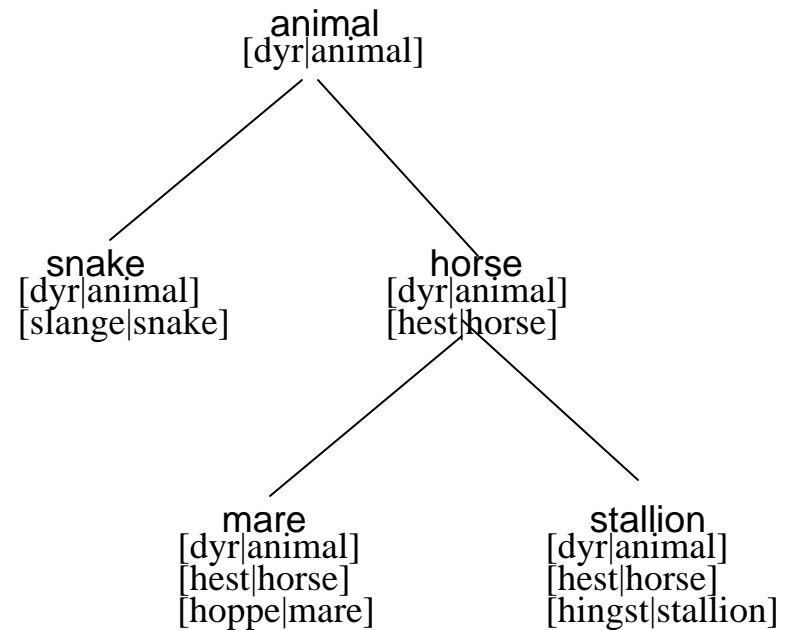
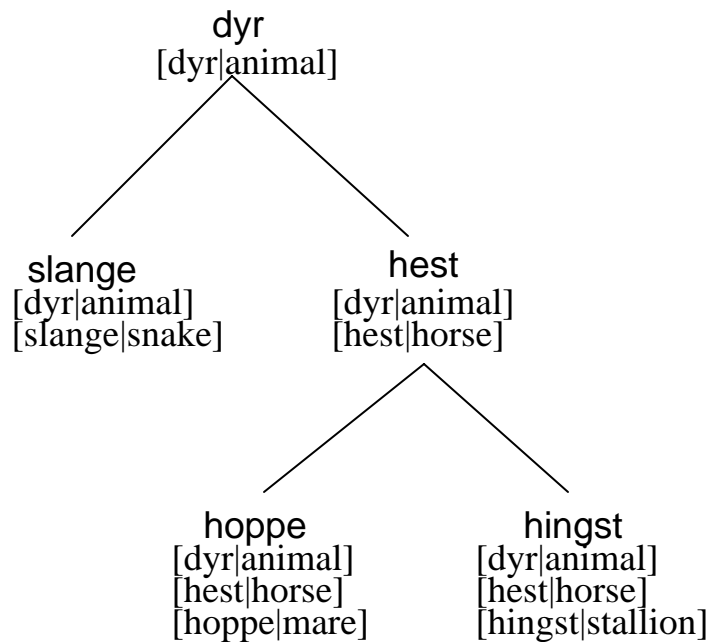


The feature sets express:

- Semantic relations among the senses
- Translational relations to the other language



Inclusion relations among the feature sets define lattices:



Automatic word alignment

Implemented by Sindre Sørensen, Aksis, Bergen

Web demo:

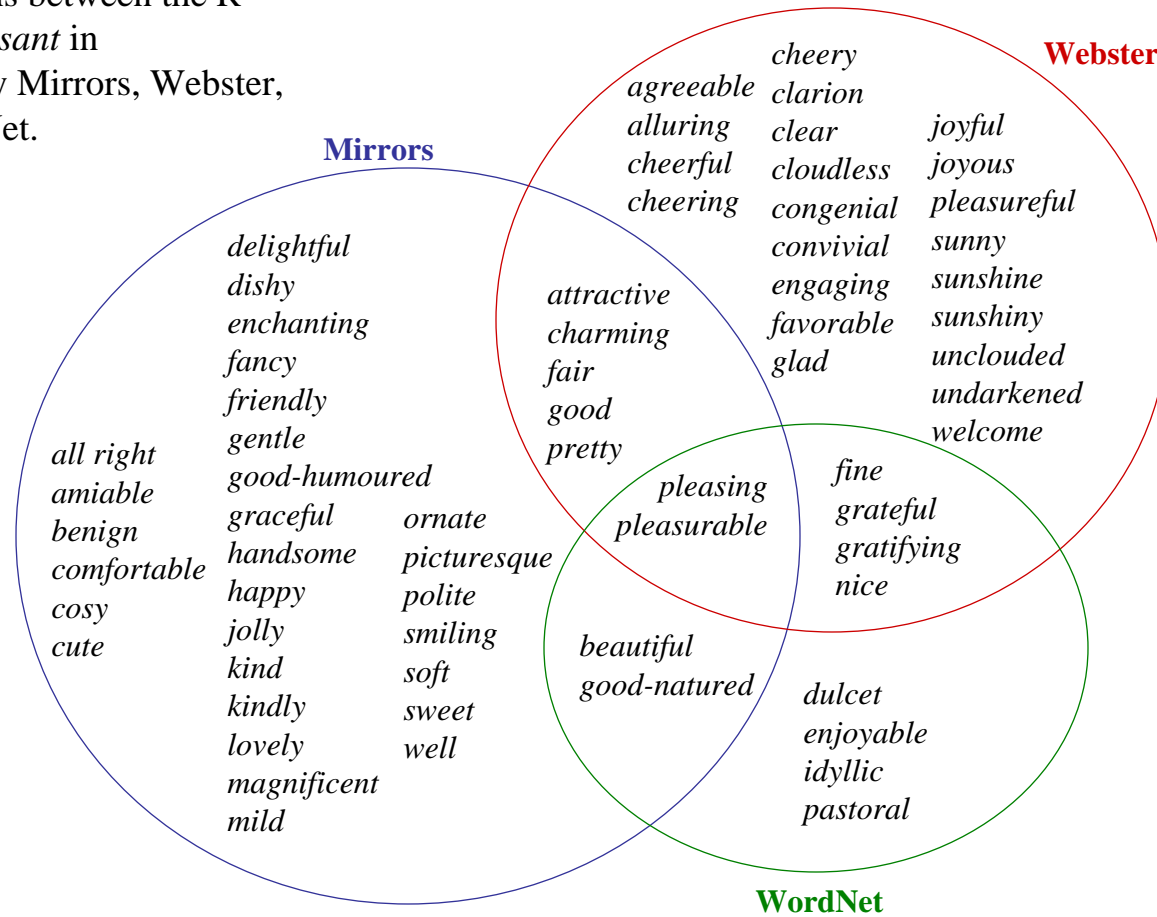
<http://decentius.hit.uib.no:9000/cl/sm/documents.html>

(Password protected)

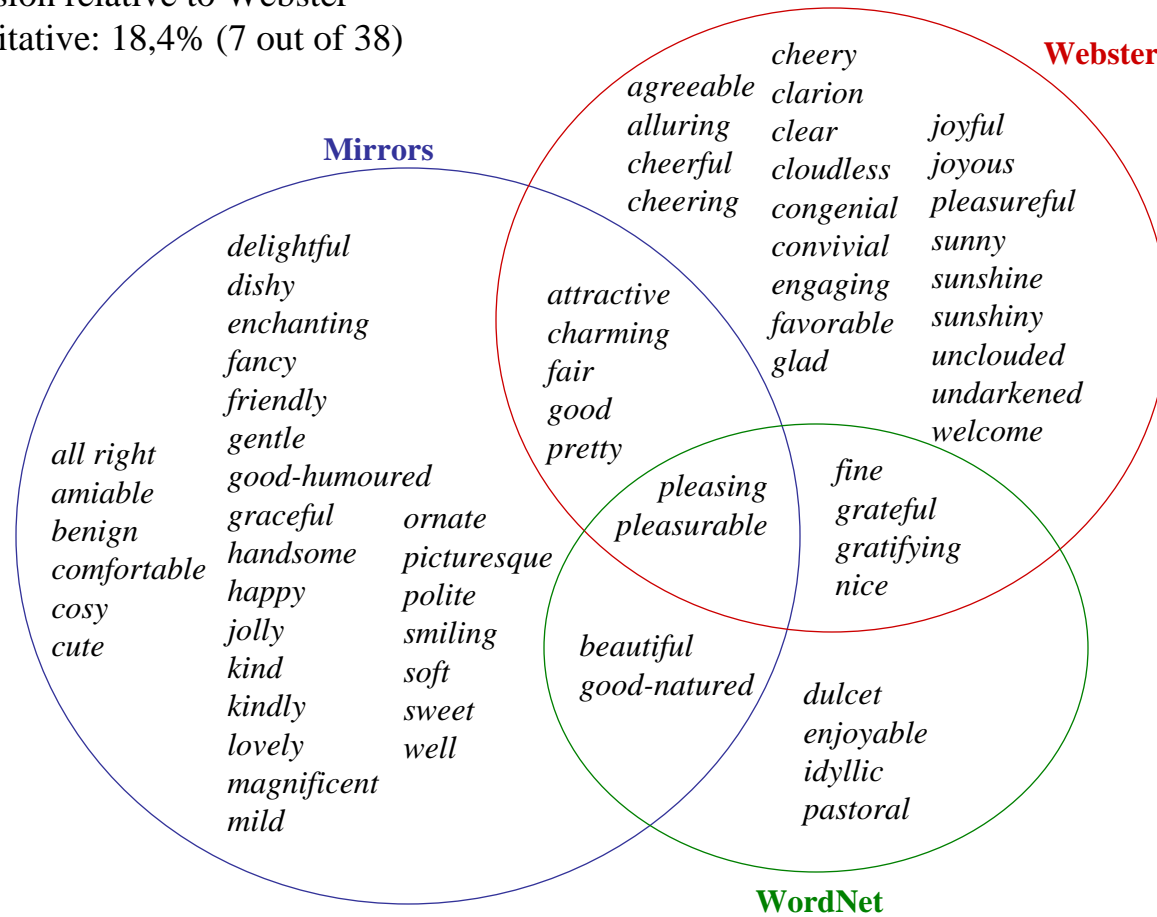
Evaluation of the Mirrors output

Work by Martha Thunes, Aksis, Bergen

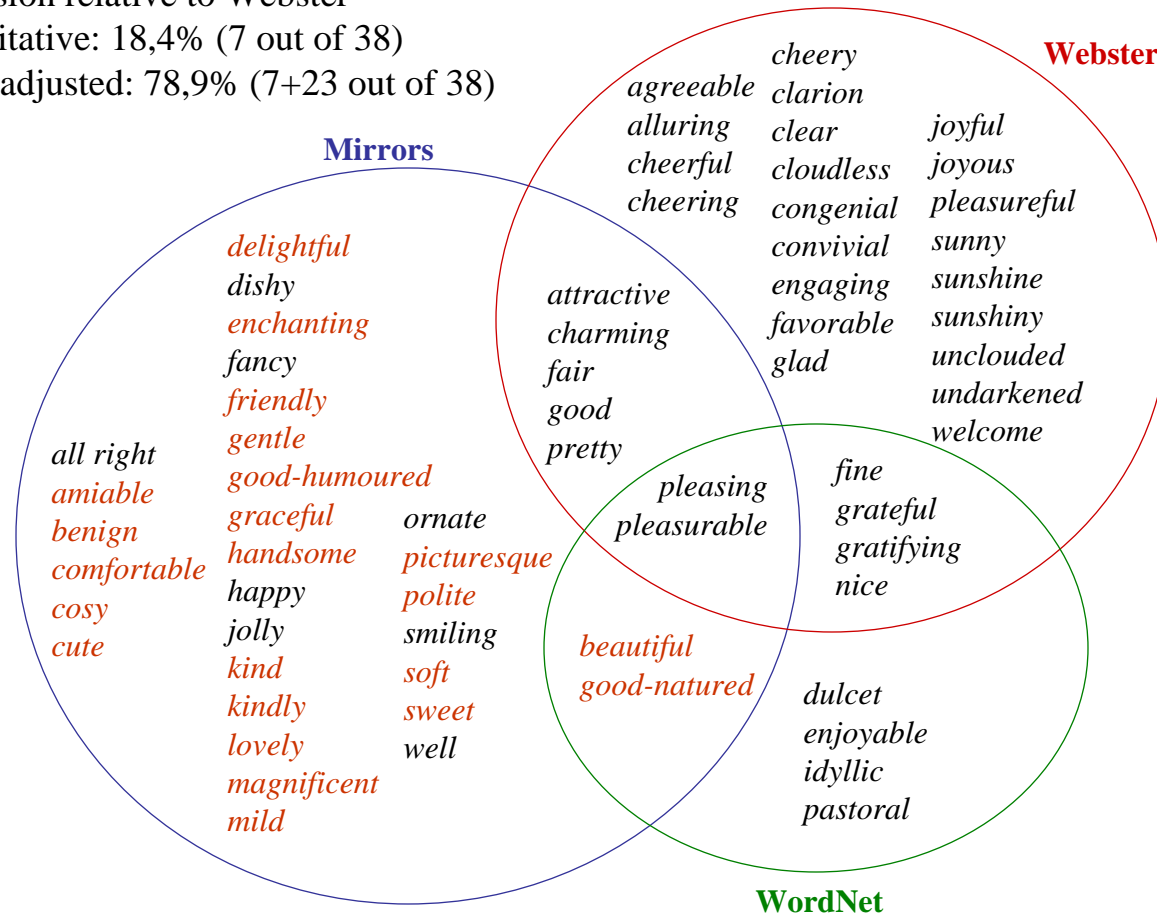
Intersections between the R-sets of *pleasant* in respectively Mirrors, Webster, and WordNet.



Mirrors precision relative to Webster
 Strictly quantitative: 18,4% (7 out of 38)

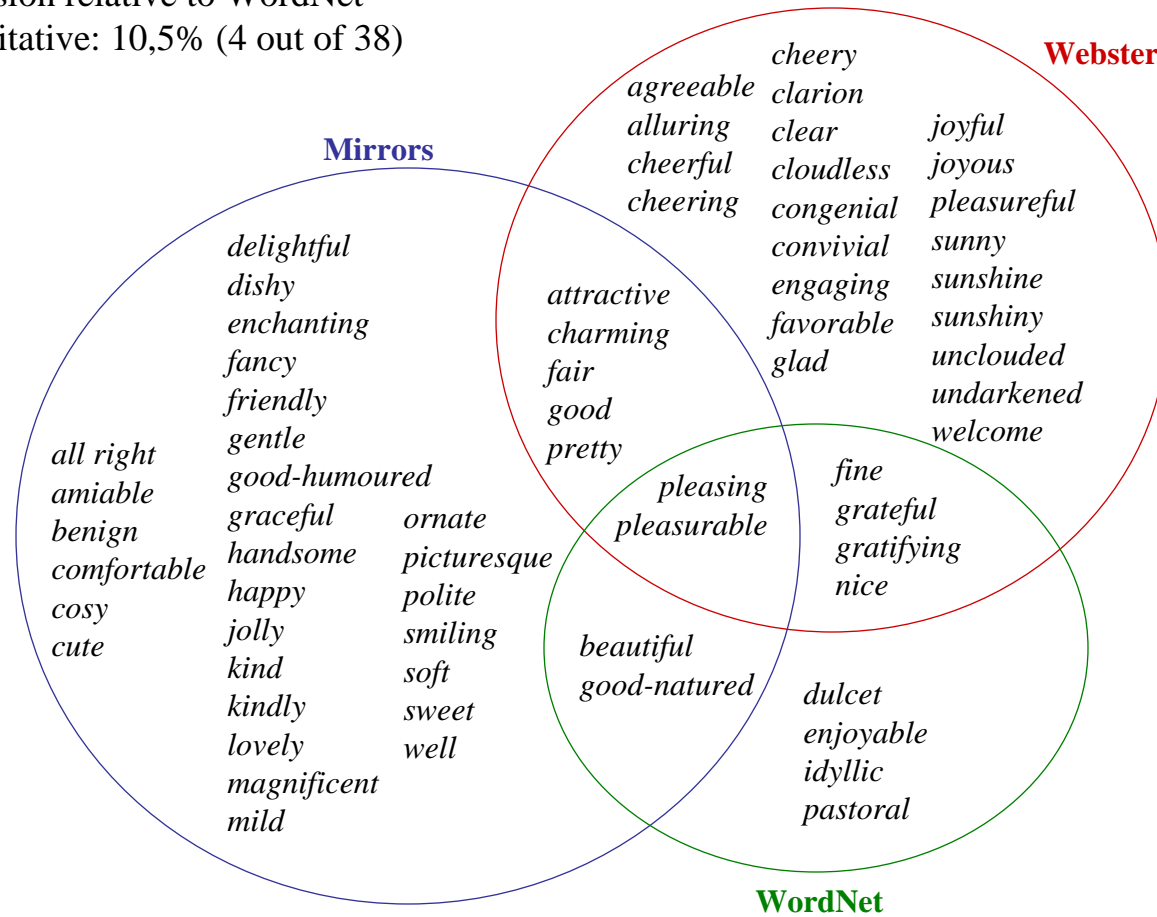


Mirrors precision relative to Webster
 Strictly quantitative: 18,4% (7 out of 38)
 Qualitatively adjusted: 78,9% (7+23 out of 38)

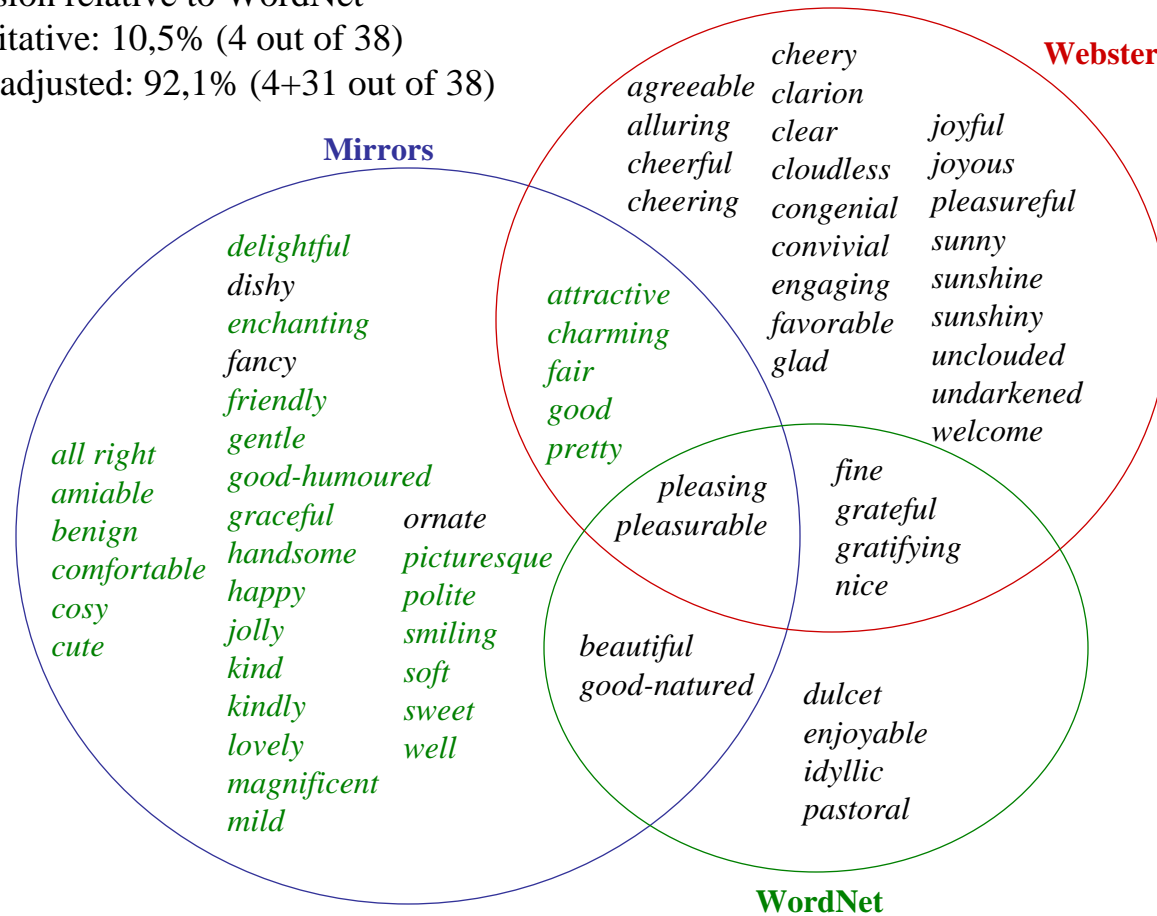


Qualitative adjustment: another 23 Mirrors R-words, in red, fall within the sense description given in the Webster entry.

Mirrors precision relative to WordNet
 Strictly quantitative: 10,5% (4 out of 38)

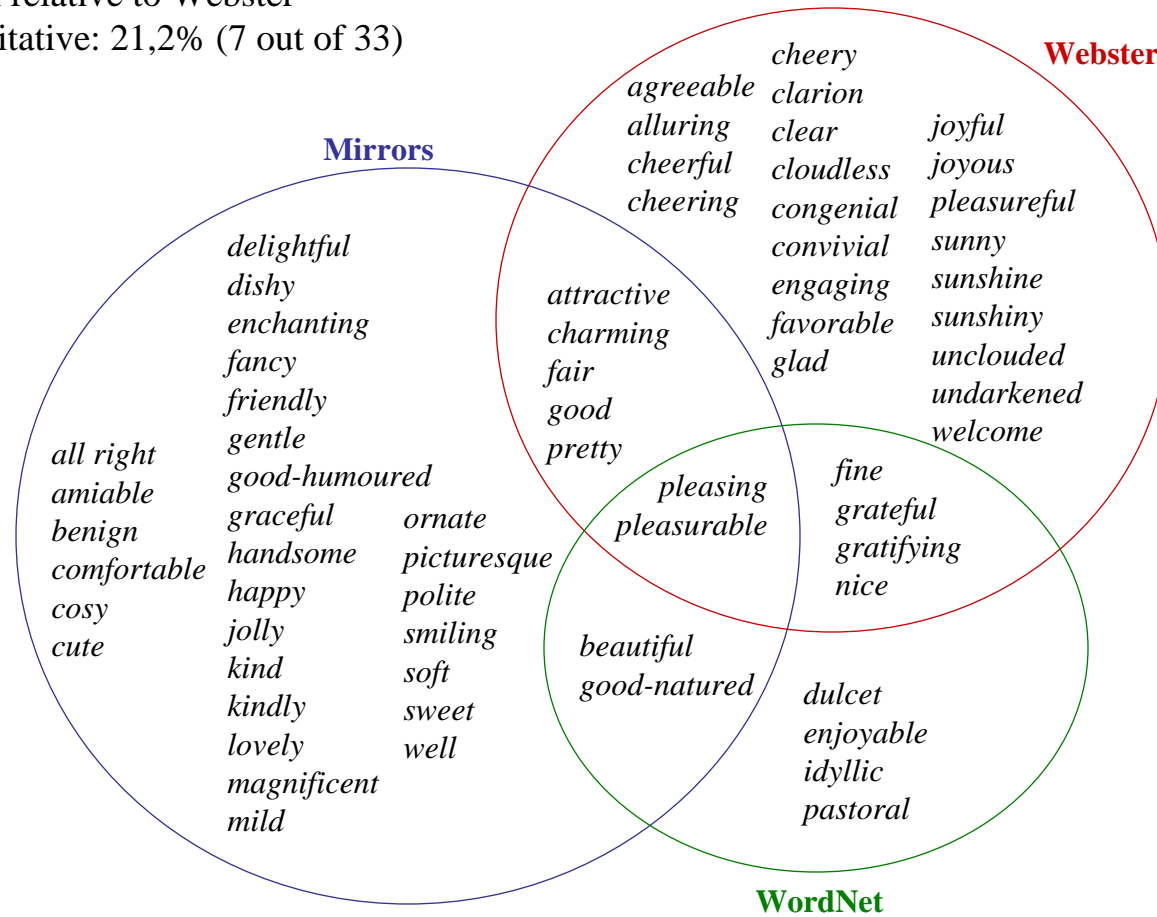


Mirrors precision relative to WordNet
 Strictly quantitative: 10,5% (4 out of 38)
 Qualitatively adjusted: 92,1% (4+31 out of 38)



Qualitative adjustment: another 31 Mirrors R-words, in green, fall within the sense description given in the WordNet entry.

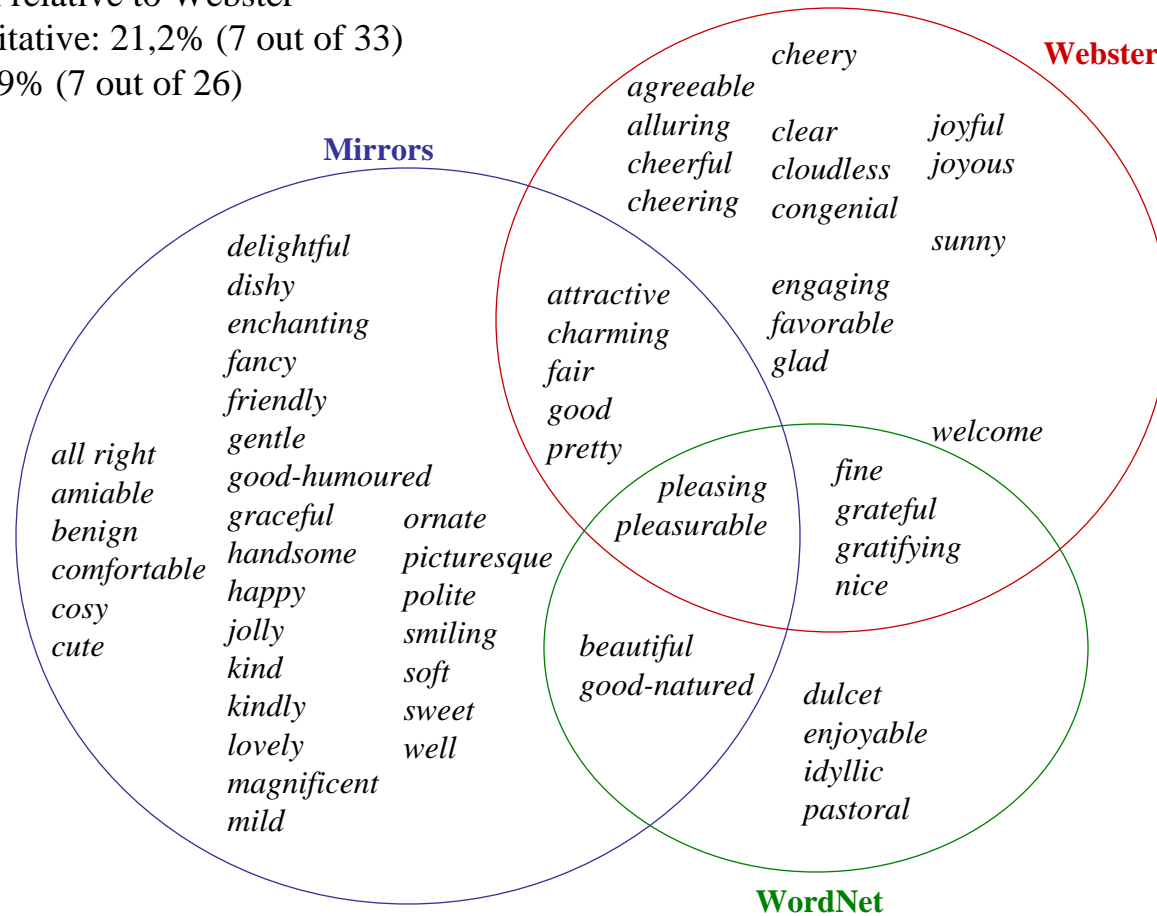
Mirrors recall relative to Webster
 Strictly quantitative: 21,2% (7 out of 33)



Mirrors recall relative to Webster

Strictly quantitative: 21,2% (7 out of 33)

Adjusted: 26,9% (7 out of 26)

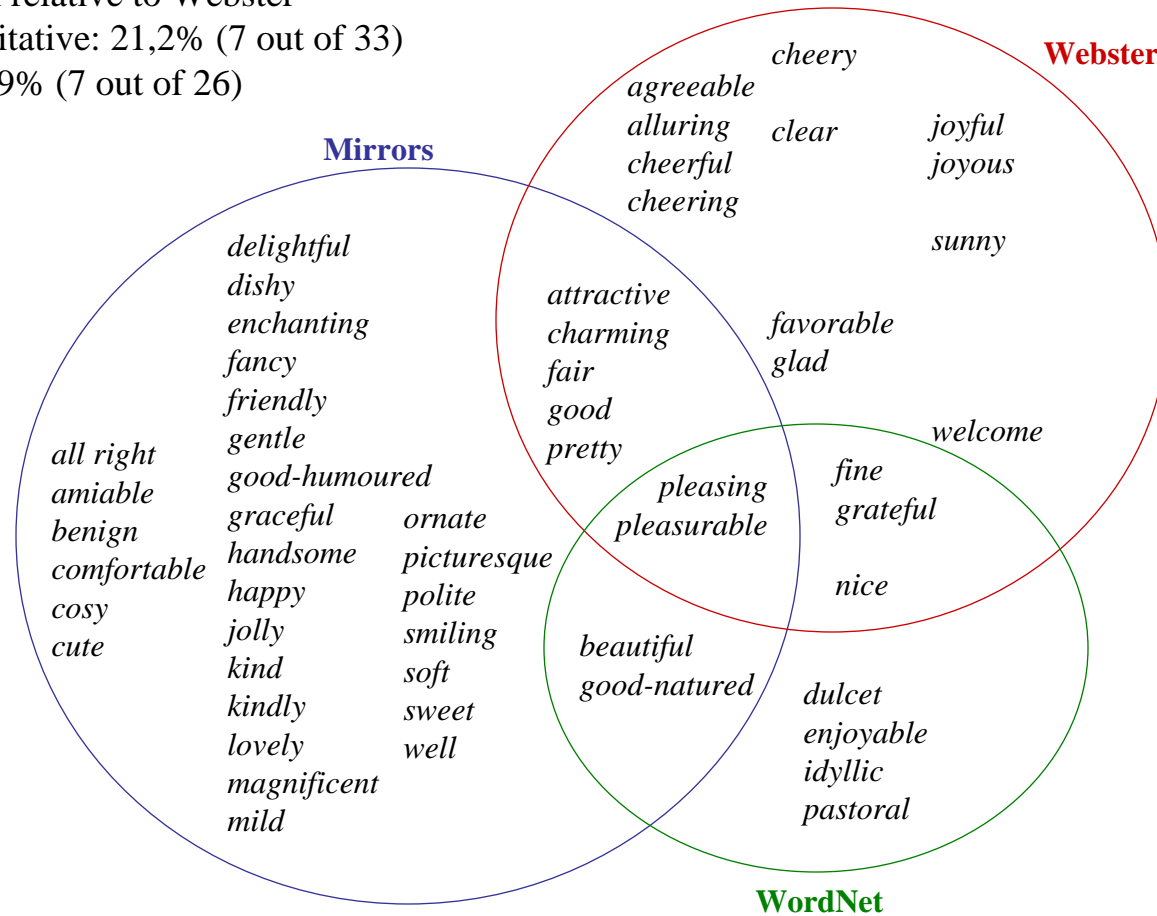


Adjusted recall: 7 Webster R-words do not occur in the corpus; hence, cannot be included in Mirrors.

Mirrors recall relative to Webster

Strictly quantitative: 21,2% (7 out of 33)

Adjusted: 26,9% (7 out of 26)

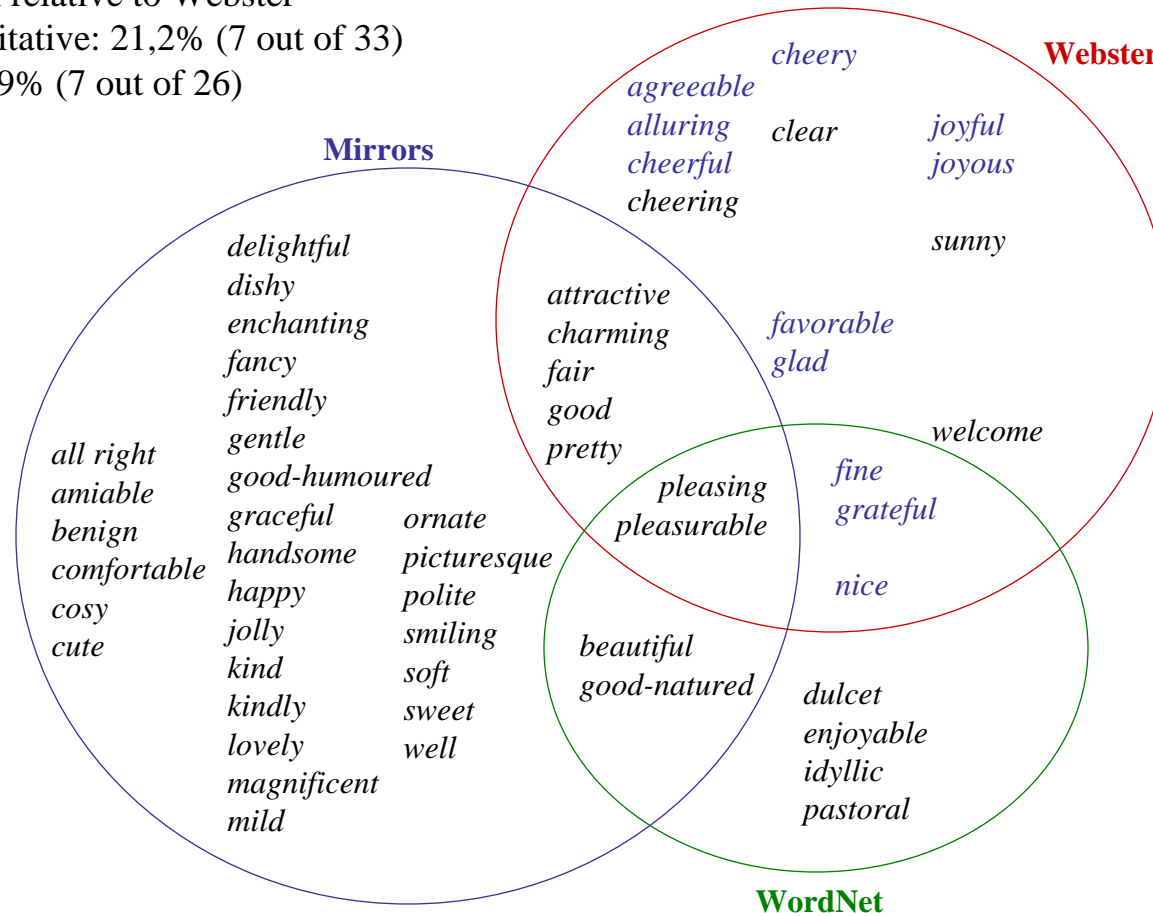


Further adjustment: disregarding another 4 Webster R-words because the Mirrors method cannot identify them as R-words due to their translational properties in the ENPC.

Mirrors recall relative to Webster

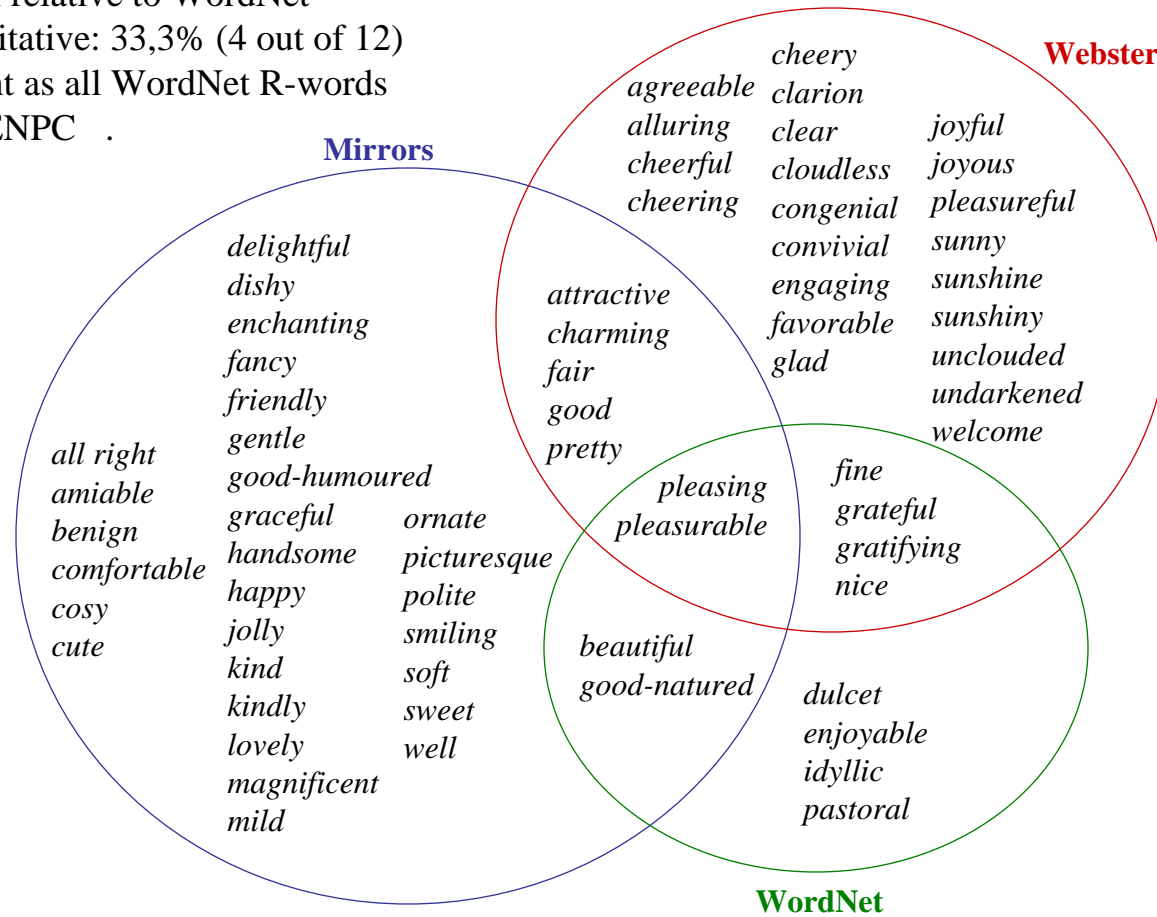
Strictly quantitative: 21,2% (7 out of 33)

Adjusted: 26,9% (7 out of 26)

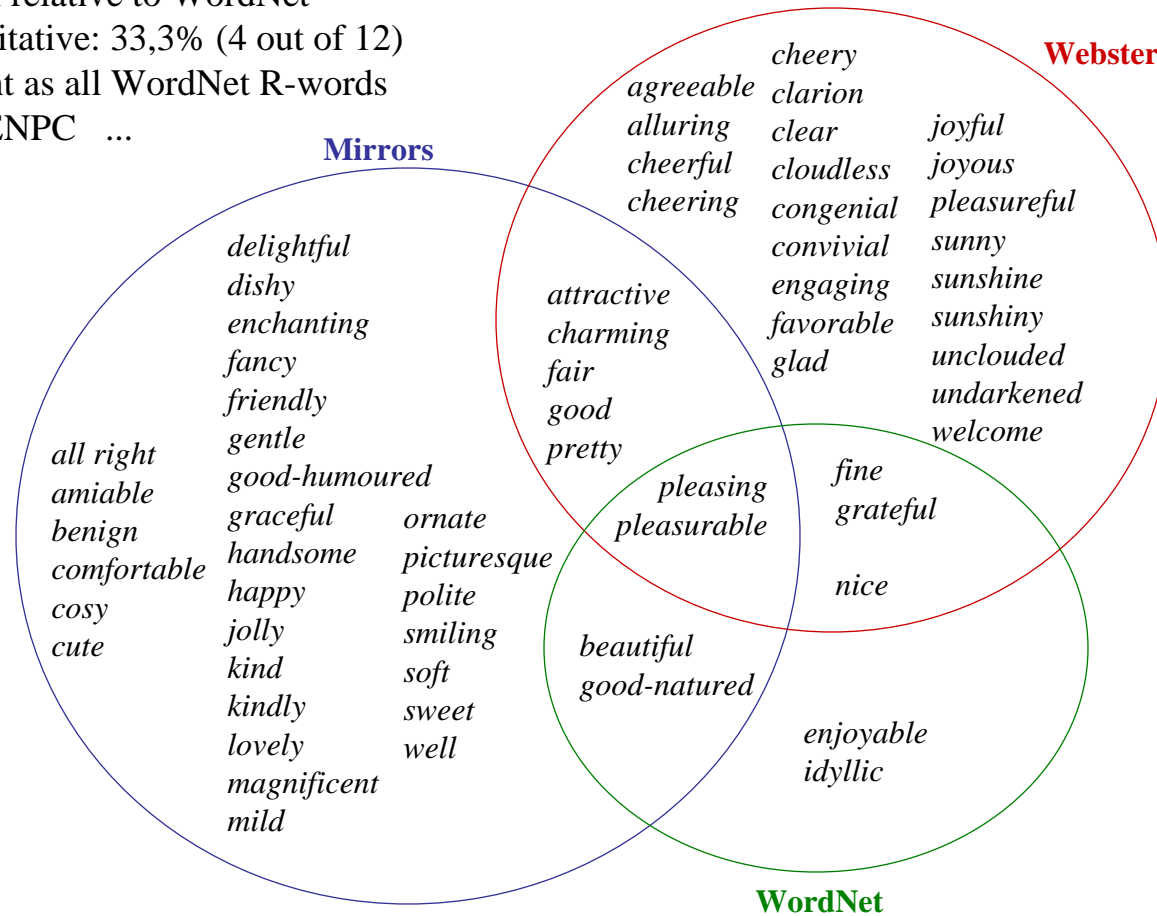


Among the remaining 15 Webster R-words missed by Mirrors, 11 words, in blue, represent meanings covered by the Mirrors R-set.

Mirrors recall relative to WordNet
 Strictly quantitative: 33,3% (4 out of 12)
 No adjustment as all WordNet R-words
 occur in the ENPC .

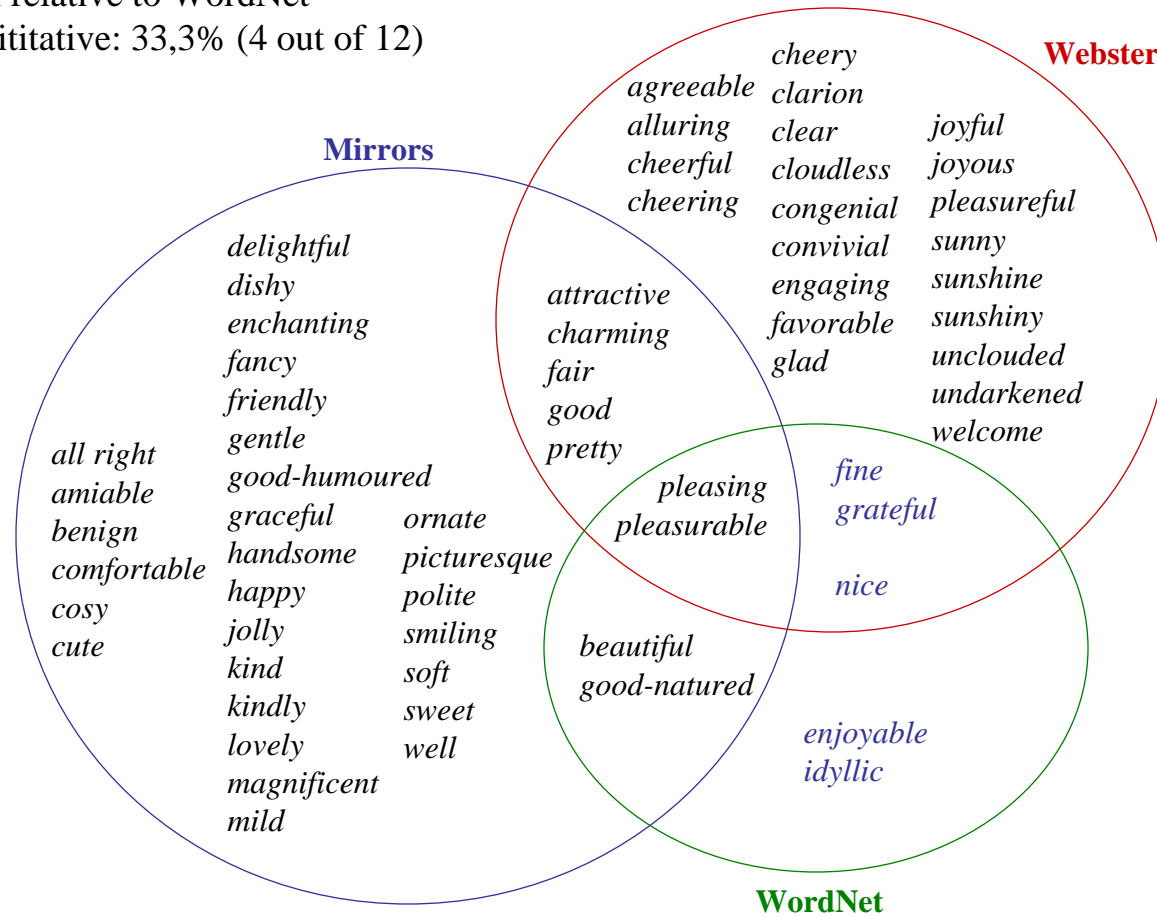


Mirrors recall relative to WordNet
 Strictly quantitative: 33,3% (4 out of 12)
 No adjustment as all WordNet R-words
 occur in the ENPC ...



...but 3 WordNet R-words may be disregarded as they are missed by Mirrors due to their translational properties in the corpus.

Mirrors recall relative to WordNet
 Strictly quantitative: 33,3% (4 out of 12)



The remaining 5 WordNet R-words missed by Mirrors represent meanings covered by the Mirrors R-set, indicated with blue print.

Entry word	No. of words in Mirrors R-set	No. of words in Webster R-set	Words shared between the two R-sets		
			in numbers	% of Mirrors total (precision)	% of Webster total (recall)
<i>able</i>	13	25	7	53,8	28,0
<i>adequate</i>	31	18	3	9,7	16,7
<i>affectionate</i>	5	9	1	20,0	11,1
<i>all right</i>	38	10	2	5,3	20,0
<i>appropriate</i>	15	40	2	13,3	5,0
<i>attractive</i>	26	47	9	34,6	19,1
<i>beneficial</i>	9	12	2	22,2	16,7
<i>bright</i>	38	66	8	21,1	12,1
<i>clear</i>	59	73	12	20,3	16,4
<i>comforting</i>	15	no entry			
<i>delicious</i>	11	20	1	9,1	5,0
<i>easy</i>	49	114	15	30,6	13,2
<i>excellent</i>	23	56	2	8,7	3,6
<i>fair</i>	52	83	13	25,0	15,7
<i>favourable</i>	5	54	4	80,0	7,4
<i>fine</i>	31	42	3	9,7	7,1
<i>firm</i>	56	47	11	19,6	23,4
<i>first-rate</i>	2	10	1	50,0	10,0
<i>fortunate</i>	8	12	2	25,0	16,7
<i>fresh</i>	53	51	9	17,0	17,6
<i>friendly</i>	28	15	0	0,0	0,0
<i>full</i>	27	44	9	33,3	20,5
<i>genuine</i>	36	26	5	13,9	19,2
<i>good</i>	112	123	26	23,2	21,1

Entry word	No. of words in Mirrors R-set	No. of words in Webster R-set	Words shared between the two R-sets		
			in numbers	% of Mirrors total (precision)	% of Webster total (recall)
<i>kind</i>	22	38	7	31,8	18,4
<i>nice</i>	39	71	3	7,7	4,2
<i>peaceful</i>	17	13	1	5,9	7,7
<i>pleasant</i>	38	33	7	18,4	21,2
<i>plentiful</i>	6	25	4	66,7	16,0
<i>positive</i>	22	56	6	27,3	10,7
<i>ripe</i>	1	21	0	0,0	0,0
<i>satisfactory</i>	27	25	5	18,5	20,0
<i>serviceable</i>	0	10	0	0,0	0,0
<i>siz(e)able</i>	2	14	1	50,0	7,1
<i>solid</i>	51	17	5	9,8	29,4
<i>sound</i>	44	56	10	22,7	17,9
<i>spectacular</i>	19	18	2	10,5	11,1
<i>steady</i>	32	41	4	12,5	9,8
<i>superb</i>	7	35	2	28,6	5,7
<i>sweet</i>	51	37	8	15,7	21,6
<i>tantalising</i>	0	no adj. entry			
<i>thorough</i>	38	15	1	2,6	6,7
<i>worthy</i>	3	21	0	0,0	0,0
SUM (minus <i>comforting</i>)	1146	1543	212	18,5	13,7

Results for 42 adjectives,
 with Merriam-Webster's Thesaurus as gold standard
 (after Thunes):

Entry word	No. of words in Mirrors R-set	No. of words in Webster R-set	Words shared between the two R-sets		
			in numbers	% of Mirrors total (precision)	% of Webster total (recall)
SUM 42 adjectives	1146	1543	212	18,5	13,7

Qualitative adjustment of precision and recall
for 3 adjectives
(after Thunes):

Precision of the Mirrors R-set of	Strictly quantitative		Qualitatively adjusted	
	in numbers	percentage	in numbers	percentage
<i>delicious</i>	1 of 11	9,1	5 of 11	45,5
<i>firm</i>	12 of 57	21,1	38 of 57	66,7
<i>fortunate</i>	2 of 8	25,0	4 of 8	50,0

Recall of the Mirrors R-set of	Strictly quantitative		Adjusted w.r.t. corpus occurrences	
	in numbers	percentage	in numbers	percentage
<i>delicious</i>	1 of 20	5,0	1 of 12	8,3
<i>firm</i>	12 of 47	25,5	12 of 43	27,9
<i>fortunate</i>	2 of 12	16,7	2 of 9	22,2

Manual vs. automatic word alignment:
 Results with Webster as gold standard
 (after Thunes):

Entry word	No. of words in Mirrors R-set	No. of words in Webster R-set	Words shared between the two R-sets		
			in numbers	% of Mirrors total (precision)	% of Webster total (recall)
SUM 42 adjectives, <i>manual alignment</i>	1146	1543	212	18,5	13,7
SUM 42 adjectives, <i>automatic alignment</i>	477	1543	90	18,9	5,8