Explorations in 3D: Semantic Tensor Spaces

Tim Van de Cruys

INRIA, ROCQUENCOURT

Distributional Semantics workshop
June 23, 2010
Groningen

Distributional similarity

Distributional similarity models are able to infer (lexical) semantics from text:

- Semantically similar words (syntactic context, small window)
  - **Parijs** 'Paris': *Londen* 'London', *Berlijn* 'Berlin', *New York*, *Milaan* 'Milan', *Rome*, *Wenen* 'Vienna', *Moskou* 'Moscow', *Stockholm*, *Frankfurt*
  - **liefde** 'love': *vriendschap* 'friendship', *passie* 'passion', *verlangen* 'desire', *angst* 'fear', *verdriet* 'sadness', *emotie* 'emotion', *schoonheid* 'beauty', *geloof* 'faith'

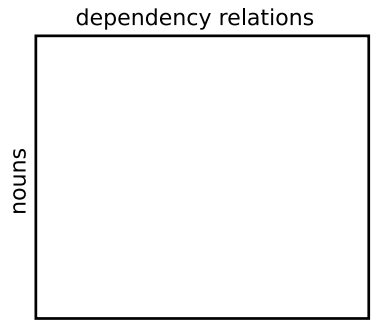
Distributional similarity

Distributional similarity models are able to infer (lexical) semantics from text:

- Semantic dimensions (large window)
  - (**transport**): *bus* 'bus', *taxi* 'taxi', *trein* 'train', *halte* 'stop', *reiziger* 'traveler', *perron* 'platform', *tram* 'tram', *station* 'station', *chauffeur* 'driver', *passagier* 'passenger'
  - (**food**): *bouillon* 'broth', *slagroom* 'cream', *ui* 'onion', *eierdooier* 'egg yolk', *laurierblad* 'bay leaf', *zout* 'salt', *deciliter* 'decilitre', *boter* 'butter', *bleekselderij* 'celery', *saus* 'sauce'

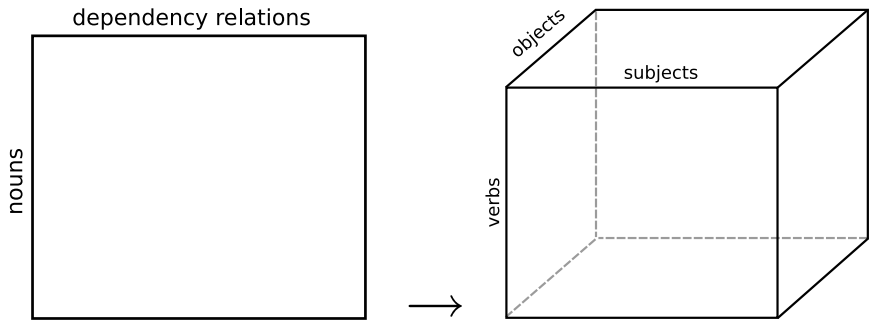
Two-way vs. three-way

- all methods use two way co-occurrence frequencies $\longrightarrow$ matrix
- suitable for two-way problems
  - words $\times$ documents
  - nouns $\times$ dependency relations
- not suitable for *n*-way problems
  - words $\times$ documents $\times$ authors
  - verbs $\times$ subjects $\times$ direct objects



Two-way vs. three-way

- all methods use two way co-occurrence frequencies $\longrightarrow$ matrix
- suitable for two-way problems
  - words $\times$ documents
  - nouns $\times$ dependency relations
- not suitable for *n*-way problems $\longrightarrow$ tensor
  - words $\times$ documents $\times$ authors
  - verbs $\times$ subjects $\times$ direct objects



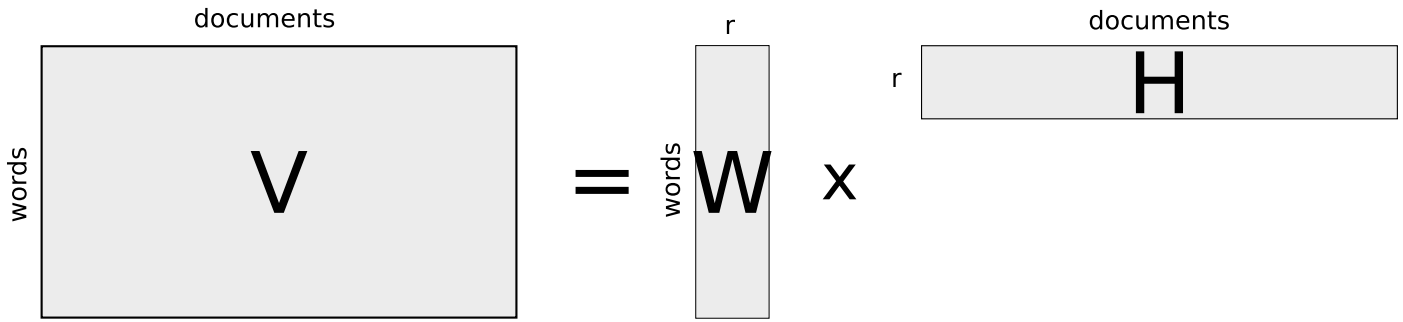
Technique

- Given a non-negative matrix V, find non-negative matrix factors W and H such that:

$$V_{nxm} \approx W_{nxr} H_{rxm} \tag{1}$$

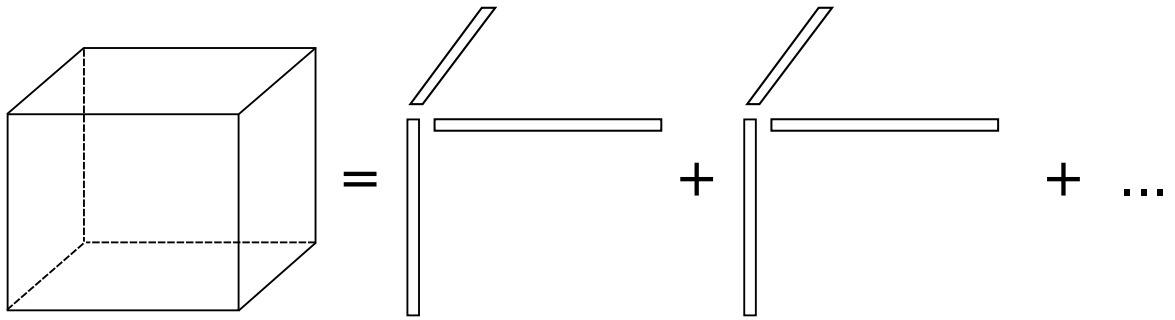
- Choosing $r \ll n, m$ reduces data
- Constraint on factorization: all values in three matrices need to be *non-negative values* (≥ 0)
- Constraint brings about a *parts-based* representation: only additive, no subtractive relations are allowed

Graphical Representation

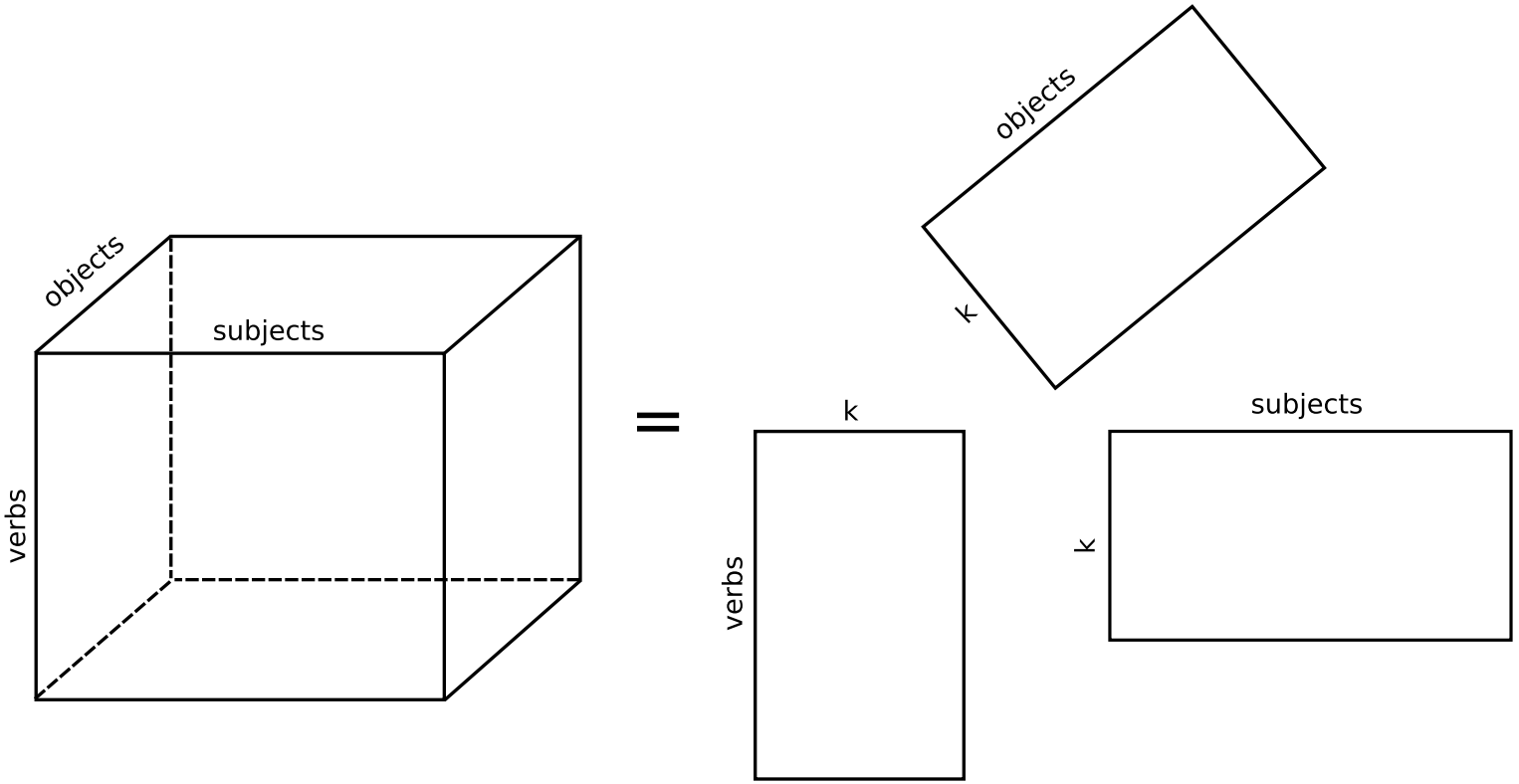


Technique

- Idea similar to non-negative matrix factorization
- Calculations are different
- $\min_{x_i \in \mathbb{R}^{D1}_{\geq 0}, y_i \in \mathbb{R}^{D2}_{\geq 0}, z_i \in \mathbb{R}^{D3}_{\geq 0}} \| T - \sum_{i=1}^{k} x_i \circ y_i \circ z_i \|_F^2$



Graphical representation



Introduction 1/2

- Standard selectional preference models: two-way co-occurrences
- Keeping track of single relationships
- But: two-way selectional preference models are not sufficiently rich
- Compare:
  - *The skyscraper is playing coffee.*
  - *The turntable is playing the piano.*

Introduction 2/2

- *The skyscraper is playing coffee.*
  - $(play, su, scyscraper)$ $\downarrow$
  - $(play, obj, coffee)$ $\downarrow$
- *The turntable is playing the piano.*
  - $(play, su, turntable)$ $\uparrow$
  - $(play, obj, piano)$ $\uparrow$
  - $(play, su, turntable, obj, piano)$ $\downarrow$

Methodology

- Three-way extraction of selectional preferences
- Approach applied to Dutch, using TWENTE NIEUWS CORPUS (500M words of newspaper texts)
- parsed with Dutch dependency parser ALPINO
- three-way co-occurrence of verbs with subjects and direct objects extracted
- adapted with extension of pointwise mutual information
- Resulting tensor 1K verbs $\times$ 10K subjects $\times$ 10K direct objects
- reduction to k dimensions ($k = 50, 100, 300$)

Examples: police action

subjects	su_s	verbs	v_s	objects	obj_s
politie 'police'	.99	*houd_aan* 'arrest'	.64	*verdachte* 'suspect'	.16
agent 'policeman'	.07	*arresteer* 'arrest'	.63	*man* 'man'	.16
autoriteit 'authority'	.05	*pak_op* 'run in'	.41	*betoger* 'demonstrator'	.14
Justitie 'Justice'	.05	*schiet_dood* 'shoot'	.08	*relschopper* 'rioter'	.13
recherche 'detective force'	.04	*verdenk* 'suspect'	.07	*raddraaier* 'instigator'	.13
marechaussee 'military police'	.04	*tref_aan* 'find'	.06	*overvaller* 'raider'	.13
justitie 'justice'	.04	*achterhaal* 'overtake'	.05	*Roemeen* 'Romanian'	.13
arrestatieteam 'special squad'	.03	*verwijder* 'remove'	.05	*actievoerder* 'campaigner'	.13
leger 'army'	.03	*zoek* 'search'	.04	*hooligan* 'hooligan'	.13
douane 'customs'	.02	*spoor_op* 'track'	.03	*Algerijn* 'Algerian'	.13

Examples: legislation

subjects	su_s	**verbs**	v_s	**objects**	obj_s
meerderheid 'majority'	.33	*steun* 'support'	.83	*motie* 'motion'	.63
VVD	.28	*dien_in* 'submit'	.44	*voorstel* 'proposal'	.53
D66	.25	*neem_aan* 'pass'	.23	*plan* 'plan'	.28
Kamermeerderheid	.25	*wijs_af* 'reject'	.17	*wetsvoorstel* 'bill'	.19
fractie 'party'	.24	*verwerp* 'reject'	.14	*hem* 'him'	.18
PvdA	.23	*vind* 'think'	.08	*kabinet* 'cabinet'	.16
CDA	.23	*aanvaard* 'accepts'	.05	*minister* 'minister'	.16
Tweede Kamer	.21	*behandel* 'treat'	.05	*beleid* 'policy'	.13
partij 'party'	.20	*doe* 'do'	.04	*kandidatuur* 'candidature'	.11
Kamer 'Chamber'	.20	*keur_goed* 'pass'	.03	*amendement* 'amendment'	.09

Examples: exhibition

subjects	su_s	**verbs**	v_s	**objects**	obj_s
tentoonstelling 'exhibition'	.50	*toon* 'display'	.72	*schilderij* 'painting'	.47
expositie 'exposition'	.49	*omvat* 'cover'	.63	*werk* 'work'	.46
galerie 'gallery'	.36	*bevat* 'contain'	.18	*tekening* 'drawing'	.36
collectie 'collection'	.29	*presenteer* 'present'	.17	*foto* 'picture'	.33
museum 'museum'	.27	*laat* 'let'	.07	*sculptuur* 'sculpture'	.25
oeuvre 'oeuvre'	.22	*koop* 'buy'	.07	*aquarel* 'aquarelle'	.20
Kunsthal	.19	*bezit* 'own'	.06	*object* 'object'	.19
kunstenaar 'artist'	.15	*zie* 'see'	.05	*beeld* 'statue'	.12
dat 'that'	.12	*koop_aan* 'acquire'	.05	*overzicht* 'overview'	.12
hij 'he'	.10	*in huis heb* 'own'	.04	*portret* 'portrait'	.11

Examples: quality count

- 44 dimensions contain similar, framelike semantics
- 43 dimensions contain less clear-cut semantics
  - single verbs account for one dimension
  - verb senses are mixed up
- 13 dimensions based on syntax rather than semantics
  - fixed expressions
  - pronomina

Evaluation: methodology

- pseudo-disambiguation task to test generalization capacity (standard automatic evaluation for selectional preferences)

s	v	o	s'	o'
jongere	*drink*	*bier*	*coalitie*	*aandeel*
'youngster'	'drink'	'beer'	'coalition'	'share'
werkgever	*riskeer*	*boete*	*doel*	*kopzorg*
'employer'	'risk'	'fine'	'goal'	'worry'
directeur	*zwaai*	*scepter*	*informateur*	*vodka*
'manager'	'sway'	'sceptre'	'informer'	'wodka'

- 10-fold cross validation ($\pm$ 300,000 co-occurrences)

Evaluation: models

- Evaluation of 4 different models
- 2 matrix models
  $\longrightarrow$ 1K verbs $\times$ (10K subjects + 10K direct objects)
  - singular value decomposition ($\mathbb{R}$)
  - non-negative matrix factorization ($\mathbb{R}_{\geq 0}$)
- 2 tensor models
  $\longrightarrow$ 1K verbs $\times$ 10K subjects $\times$ 10K direct objects
  - parallel factor analysis ($\mathbb{R}$)
  - non-negative tensor factorization ($\mathbb{R}_{\geq 0}$)

Evaluation: results

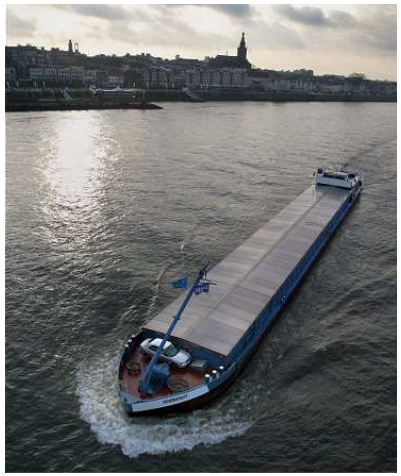
	dimensions		
	50 (%)	100 (%)	300 (%)
SVD	69.60 ± 0.41	62.84 ± 1.30	45.22 ± 1.01
NMF	81.79 ± 0.15	78.83 ± 0.40	75.74 ± 0.63
PARAFAC	85.57 ± 0.25	83.58 ± 0.59	80.12 ± 0.76
NTF	**89.52** ± 0.18	**90.43** ± 0.14	**90.89** ± 0.16

Introduction

- **Problem**: ambiguity
- WAAL

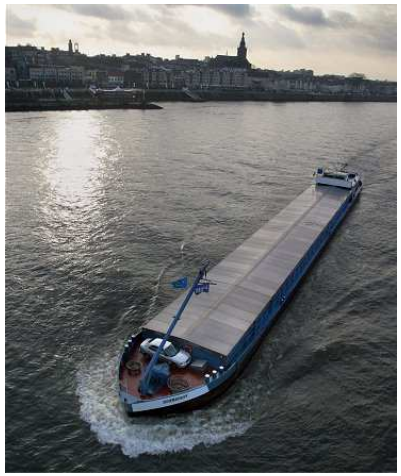
Introduction

- **Problem**: ambiguity
- WAAL



Introduction

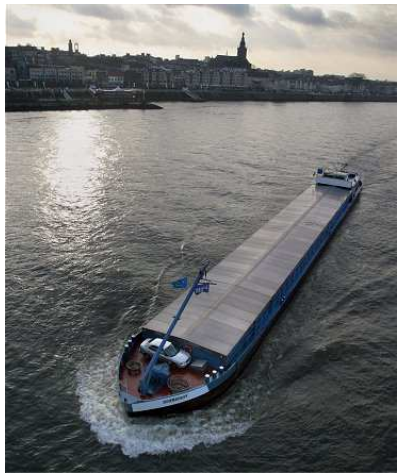
- **Problem**: ambiguity
- WAAL



$\longleftrightarrow$

Introduction

- **Problem**: ambiguity
- WAAL



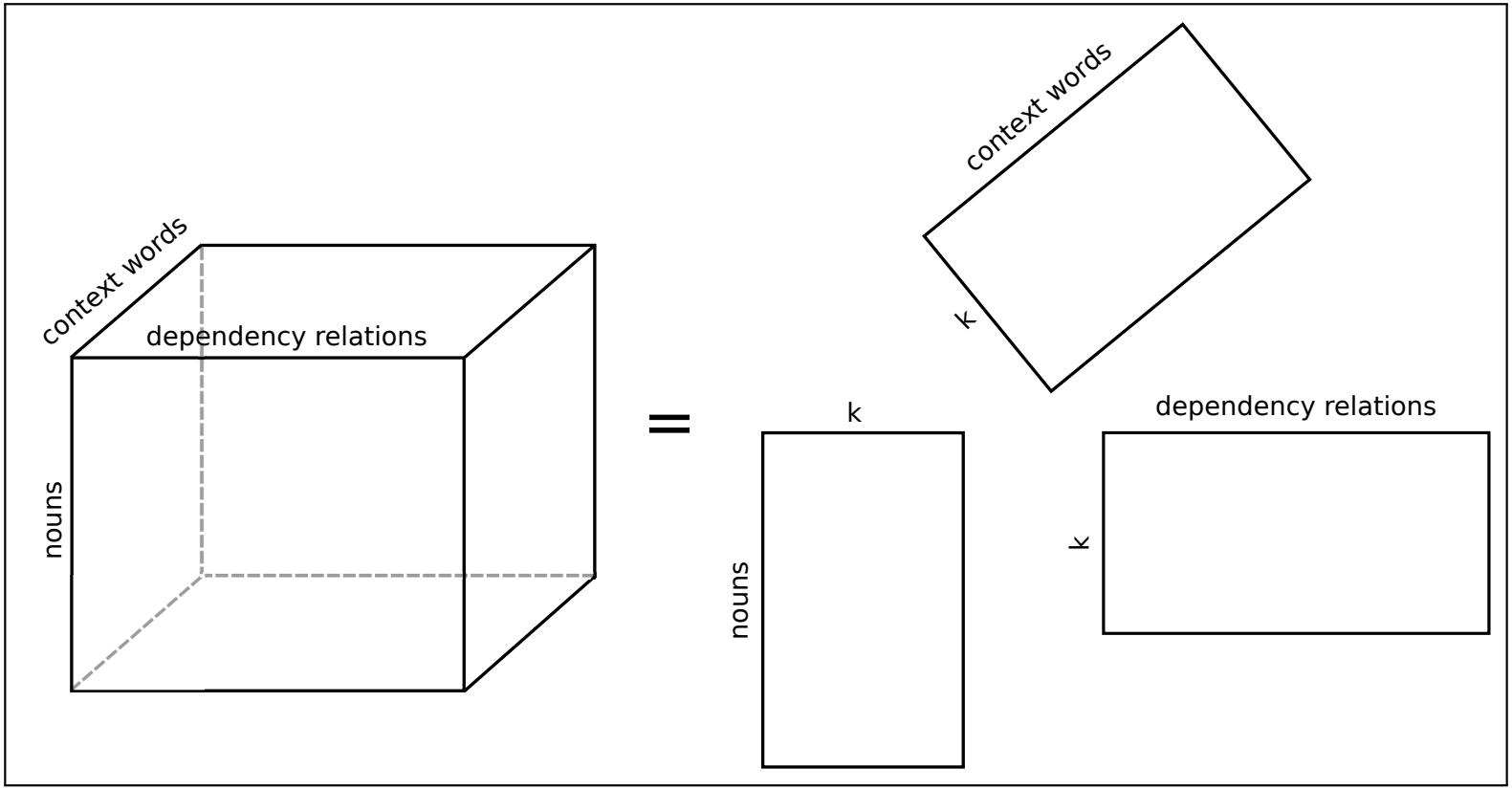
$\longleftrightarrow$

- Main research question: can 'bag of words' context and syntactic context be combined to differentiate between various senses of a word?

Methodology

- Goal: classification of nouns according to both 'bag of words' context and syntactic context
- $\Rightarrow$ extract three-way co-occurrences of nouns with dependency relations with bag of words context words
- $\Rightarrow$ three-way tensor of $5k$ nouns $\times$ $80k$ dependency relations $\times$ $1k$ context words
- Factorized with non-negative tensor factorization to 50 dimensions
- (For Dutch, TWNC, parsed with Alpino, paragraph as context)

Graphical Representation



Sense subtraction

- 'switch off' one dimension of an ambiguous word to reveal other possible senses
- From matrix W, we know which dimensions are the most important for a certain word
- Matrix H gives the importance of each dependency relation given a dimension
- 'subtract' dependency relations that are responsible for a given dimension from the original noun vector
  - $\overrightarrow{v}_{new} = \overrightarrow{v}_{orig}(\overrightarrow{1} - \overrightarrow{h}_{dim})$
  - each dependency relation is multiplied by a scaling factor, according to the load of the feature on the subtracted dimensions

Example dimension: food

nouns	**dependency rels**	**context words**
knoflook 'garlic'	$voeg\_toe_{OBJ}$ 'add'	*voeg_toe* 'add'
peper 'pepper'	$roer_{OBJ}$ 'stir'	*minuut* 'minute'
zout 'salt'	$zout_{COO}$ 'salt'	*vuur* 'fire'
ui 'oignon'	$peper_{COO}$ 'pepper'	*water* 'water'
peterselie 'persil'	$knoflook_{COO}$ 'garlic'	*stuk* 'piece'
tomaat 'tomato'	$meng_{OBJ}$ 'mix'	*leg* 'lay'
gember 'ginger'	ui_{COO} 'oignon'	*rood* 'red'
suiker 'sugar'	bak_{OBJ} 'cook'	*vis* 'fish'
citroen_sap 'lemon juice'	$strooi_{OBJ}$ 'sprinkle'	*dik* 'thick'
koriander 'coriander'	$laurier_{COO}$ 'bay leaf'	*warm* 'warm'

Example dimension: music

nouns	**dependency rels**	**context words**
muziek 'music'	$klink_{SUB}$ 'sound'	*muziek* 'music'
dans 'dance'	$klassiek_{ADJ}$ 'classical'	*dans* 'dance'
jazz	$componeer_{OBJ}$ 'compose'	*klassiek* 'classical'
liedje 'song'	$dans_{COO}$ 'dance'	*cd*
nummer 'song'	$zing_{OBJ}$ 'sing'	*lied* 'song'
lied 'song'	$hoor_{OBJ}$ 'hear'	*klink* 'sound'
cd	$draai_{OBJ}$ 'play'	*zing* 'sing'
melodie 'melody'	$theater_{COO}$ 'theatre'	*theater* 'theatre'
opera 'opera'	$film_{COO}$ 'film'	*nummer* 'song'
rock	$literatuur_{COO}$ 'literature'	*tekst* 'text'

Conclusion

- Tensor space: novel model to investigate three-way (n-way) co-occurrence data
- Possible to generalize over co-occurrence data with appropriate factorization models
- Model applicable to and beneficial for:
  - Three-way selectional preference induction
  - Word sense discrimination