

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



“Vreselijk mooi!” (terribly beautiful): A Subjectivity Lexicon for Dutch Adjectives

- 3,000 Dutch adjectives tagged with:

- polarity **-1.0** → **+1.0**

- subjectivity 0.0 → 1.0

- intensity 0.5 → 2.0

- word sense

- CORNETTO id

- WORDNET id

belabberd -0.6

uitstekend +0.9

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



“Vreselijk mooi!” (terribly beautiful): A Subjectivity Lexicon for Dutch Adjectives

- XML formatted
- Open source (bundled in PATTERN)
- F-score: ~0.75
- Compact: 750KB



Sentiment Analysis **WHAT** **HOW**

- Discern facts from opinions in text
- Polarity: positive or negative opinion?
- Target-holder identification

“I bought an iPhone a few days ago.¹ It was such a nice phone.² The touch screen was really cool.³ Although the battery life was not long, that is ok for me.⁴ However, my mother was mad with me as I did not tell her before I bought it.⁵”

2, 3	positive opinion
4, 5	negative opinion
2, 3, 4	target: iPhone (touch screen / battery life)
5	target: me
2, 3, 4	holder: me
5	holder: mother



Sentiment Analysis

WHAT

HOW

- Subjectivity lexicon
 - dictionary of words with positive / negative scores (polarity)
 - Taboada et al., 2011
- Machine learning text classification
 - Pang et al., 2002

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



Subjectivity lexicon of adjectives

WHY

HOW

Adjectives and adverbs are classified more frequently as subjective (40% and 35%) than verbs (10%).

- Esuli & Sebastiani, 2006

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans

Subjectivity lexicon of adjectives

WHY

HOW

- Which adjectives?
 - Subjective adjectives will occur frequently in product reviews
 - Collect Dutch book reviews



star rating

review

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



Subjectivity lexicon of adjectives

WHY

HOW

- Which adjectives?
 - 10,000+ book reviews from bol.com mined with PATTERN
 - Part-of-speech tagged with FROG

1	spannend	spannend	[spannend]	ADJ(vrij,basis,zonder)
2	en	en	[en]	VG(neven)
3	onvoorspelbaar	onvoorspelbaar	[on][voor][spel][baar]	ADJ(vrij,basis,zonder)
4	.	.	[.]	LET()

- 4,200 unique adjective forms
- Take **top 1,000 most frequent**
(goed, echt, heel, ...)

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans

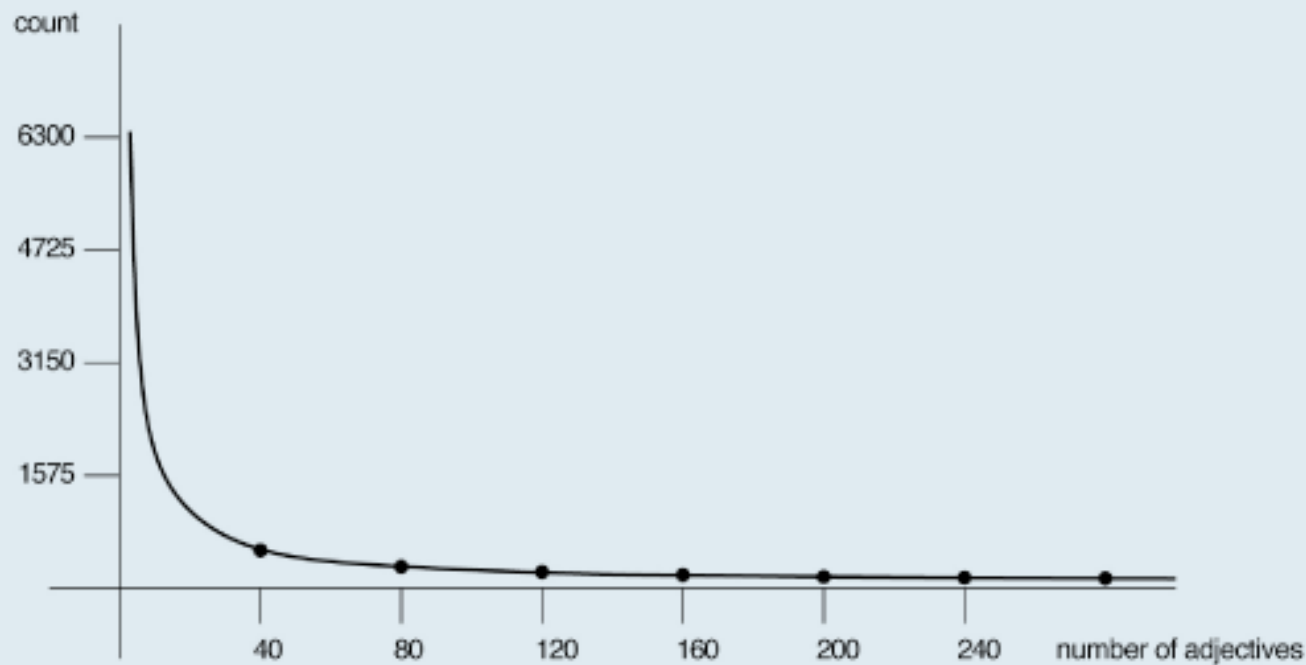


Subjectivity lexicon of adjectives

WHY

HOW

- Top 1,000



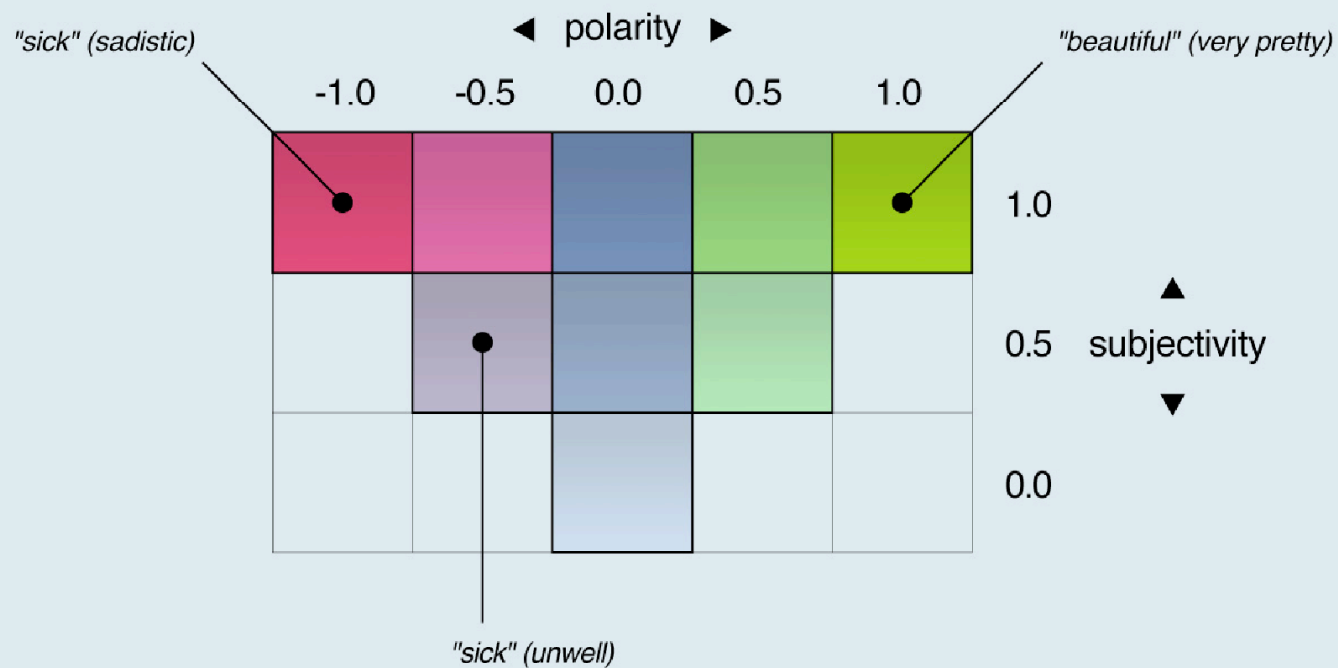
Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



Manual annotation

- Top 1,000
- 7 annotators



Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans

Manual annotation

- Top 1,000
- 7 annotators

http://www.clips.ua.ac.be/experiments/adjectieven/

Vb. van orde: vreselijk → slecht → middelmatig → goed → geweldig

[42] verrassend

Zoals in: *onverwacht*

Cometto-ID: r_a-16005
Frequentie: 62

*	**	***	****	*****
0.00	0.00	0.16	0.51	0.33

ALS LOSSTAAND ADJECTIEF

negatief ← Subjectief → positief

Objectief

ALS MODIFIER

Versterkt +

Zwakt af -

0.5/0.5

2.0

volgende

polarity / subjectivity

intensity



Manual annotation

- Top 1,000
 - Polarity
 - Subjectivity
 - Intensity

ongelooflijk goed = *incredibly good*
≠ *unbelievable + good*

x2 modifier



Manual annotation

- Top 1,000
- 7 annotators

Inter-annotator agreement:

RATING	<i>K</i>
polarity (-1 or 0 or +1)	+0.63
polarity (-1.0 => +1.0)	+0.47
polarity + subjectivity	+0.30
subjectivity	+0.34
intensity	+0.32

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



Automatic expansion **ONE** **TWO**

- Distributional expansion
 - Van de Cruys, 2010
 - Adjective vectors with nouns as features:

VECTOR	FEATURES	
	birthday	slime
<i>happy</i>	4	0
<i>great</i>	3	0
<i>disgusting</i>	0	1
<i>oozing</i>	0	3

- Nearest-neighbor yields sets of related adjectives:
happy + great | *oozing + disgusting* | ...

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



Automatic expansion

ONE

TWO

- Distributional expansion
 - 3M words from TwNC
 - 2,500 most frequent nouns
 - 6,000 adjectives preceding these nouns
 - Use **cosine similarity** to find nearest-neighbors of our 1,000 hand-tagged adjectives.

fantastisch → *geweldig, prachtig, uitstekend, prima, mooi, goed*

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



Automatic expansion

ONE

TWO

- Distributional expansion
 - 3M words from TwNC
 - 2,500 most frequent nouns
 - 6,000 adjectives preceding these nouns
 - Use **cosine similarity** to find nearest-neighbors of our 1,000 hand-tagged adjectives.

fantastisch → *geweldig, prachtig, uitstekend, prima, mooi, goed*

+1.0



Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



Automatic expansion

ONE

TWO

- Distributional expansion
 - 3M words from TwNC
 - 2,500 most frequent nouns
 - 6,000 adjectives preceding these nouns
 - Use **cosine similarity** to find nearest-neighbors of our 1,000 hand-tagged adjectives.

fantastisch → *geweldig, prachtig, uitstekend, prima, mooi, goed*

+1.0

lieflijk

+0.4

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans

Automatic expansion

ONE

TWO

- Distributional expansion
- Filter out antonyms and noise

http://www.clips.ua.ac.be/experiments/adjectieven/relaties/1/index.php

1098 **verrukkelijk** (zalig) 93% POSITIEF 100% SUBJECTIEF Next

Welke woorden drukken hetzelfde gevoel uit? (positief-negatief, subjectief-objectief)

heerlijk (heel lekker smakend)	heerlijk (heel prettig)	v.d. heer	prachtig (heel goed)	heerlijk (heel mooi)	voortreffelijk (uitnemend)	pretentius (aanmatigend)
kostelijk (uitstekend)	lekker (van smaak)	lekker (van geur)	vermakelijk (amusaant)	larmoyant (huilerig)	onderhoudend (boeiend)	langdradig (wijdlopieg)
apocrief (niet officieel)	schitterend (prachtig)	glinsterend	autobiografisch (het leven van de schrijver betreffend)	ontroerend (wat ontroert)		
spannend (enerverend)	strak	leerzaam (instructief)	amusant (vermakelijk)	aanbevelenswaardig (aanbeveling verdienen)	instructief (leerzaam)	
smakelijk (lekker)	smeug					

volgende: griezelig (eng) -79% 100%

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



Automatic expansion

ONE

TWO

- Traverse CORNETTO relations
 - is-synonym-of
 - is-near-synonym
 - is-antonym-of



Evaluation

- 2,000 Dutch book reviews + star rating
 - 1,000 negative



- 1,000 positive



Split review into words

Calculate average polarity score ($> +0.1$ = positive)

Compare to star rating



Evaluation

- 2,000 Dutch book reviews + star rating

“Eerst heb ik 'Komt een vrouw bij de dokter' gelezen en omdat ik het erg leuk vond heb ik het vervolg besteld. Ik vond het eerste boek fantastisch, maar de 2e boek vond ik ook erg leuk en boeiend om te lezen hoe het leven van de weduwnaar met zijn dochter verder gaat.”



POSITIVE



Evaluation

- 2,000 Dutch book reviews + star rating

“Eerst heb ik 'Komt een vrouw bij de dokter' gelezen en omdat ik het erg leuk vond heb ik het vervolg besteld. Ik vond het eerste boek fantastisch, maar de 2e boek vond ik ook erg leuk en boeiend om te lezen hoe het leven van de weduwnaar met zijn dochter verder gaat.”



POSITIVE

erg leuk	0.6 x 2
fantastisch	0.5
erg leuk	0.6 x 2
boeiend	0.9
verder	0.0
AVERAGE	0.76

POSITIVE

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



Evaluation

LEXICON		2,000 BOOK REVIEWS				2,000 MOVIE REVIEWS			
		A	P	R	F1	A	P	R	F1
manual	1,000	.75	.72	.82	.77	.72	.69	.78	.73
expanded 1	3,000	.75	.72	.82	.77	.72	.70	.77	.73
expanded 2	5,500	.74	.72	.79	.75	.71	.70	.71	.71

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



Evaluation

- With trivial detection of negation (*niet, nooit, geen*):

LEXICON		2,000 BOOK REVIEWS				2,000 MOVIE REVIEWS			
		A	P	R	F1	A	P	R	F1
manual	1,000	.79	.77	.83	.80	.76	.74	.79	.76
expanded 1	3,000	.79	.77	.83	.80	.76	.75	.78	.76
expanded 2	5,500	.78	.77	.80	.78	.75	.75	.72	.74

Subjectivity Lexicon for Dutch Adjectives

CLiPS, Tom De Smedt & Walter Daelemans



English translation

- From CORNETTO word id to WORDNET synset id
- Manual expansion with top 500 frequent English adjectives

LEXICON		2,000 MOVIE REVIEWS			
		A	P	R	F1
english	1,250	.72	.72	.71	.72