

Method for Building a Multidimensional Affect Dictionary for a New Language Semi-automatically

Guillaume Pitel
Gregory Grefenstette
CEA LIST, France

LREC 2008
Marrakesh, Morocco

Contacts: guillaume.pitel@gmail.com,
gregory.grefenstette@cea.fr

Acknowledgments : Fondation Lagardère, ARC RAPSODIS
(LORIA-INRIA Grand Est)

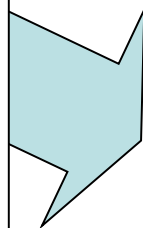


Emotive Level of Text

Maybe it was the unfriendly attitude of those hanging around the old complex.

Things started to make sense in November 2000, when authorities raided the site -- and said they found enough chemicals to make millions of doses of LSD.

"My husband and I started asking ourselves why they were working in the middle of the night. We thought it was pretty strange," said Lori Morrissey, who lives adjacent to the fenced, 26-acre site in a rural area slowly being overtaken by homes and families.



Entities: Lori Morrissey, November 2000, LSD

Content: complex, chemicals, Doses, site, husband

Emotive: unfriendly, strange, raided

Stopwords: it, the, of, to, and, a, by,....



Affect Lexicons for English

1. Lasswell Value Dictionary (1969)

✓ Eight dimensions:

- WEALTH, POWER, RECTITUDE, RESPECT, ENLIGHTENMENT, SKILL, AFFECTION, AND WELLBEING with *positive* or *negative* orientation
- e.g., *admire*: RESPECT (*positive*)

2. General Inquirer dictionary (Stone, *et al.* 1965) 9051 headwords

- 1,915 positive and 2,291 negative words (Pos/Neg)
- also labels: Active, Passive, ... , Pleasure, Pain, ... Human, Animate, ..., Region, Route, ..., Fetch, Stay, ..

<http://www.wjh.harvard.edu/~inquirer/inqdict.txt>



Clairvoyance Affect Lexicon

<lexical entry> <POS> <class> <centrality> <intensity>

"arrogance"	sn	"superiority"	0.7	0.9
..				
"gleeful"	adj	"happiness"	0.7	0.6
"gleeful"	adj	"excitement"	0.3	0.6
...				

42 pair affect classes (positive/negative)

http://www.infonortics.com/searchengines/sh01/slides-01/evans_files/v3_document.htm

Building an Affect Lexicon for a new Language

- 1. Define Affect dimensions (manual step)**
 - ✓ *3 hours*
- 2. Choose a small set of Seed Words for each dimension endpoint (manual)**
 - ✓ *One day*
 - ✓ *We chose two sizes of « small »: 2-5 or 10*
- 3. (For testing: create Gold Standard)**
 - ✓ *~5000 word-to-class mappings: 2 weeks*
 - ✓ *Only 1 native speaker*
- 4. Discover possible affect words (automatic)**
- 5. Place candidates along axes (automatic)**



Defining 44 Affect Axes for French

Axis#	Positive class	Negative class	Axis#	Positive class	Negative class
01	Avantage (51)	Désavantage (28)	23	Facilitation (92)	Obstruction (83)
02	Amour (65)	Haine (44)	24	Bienfait (54)	Crime (200)
03	Entente (27)	Opposition (95)	25	Joie (130)	Tristesse (121)
04	Fidélité (28)	Traîtrise (36)	26	Bon sens (70)	Absurde (68)
05	Attraction (25)	Répulsion (18)	27	Santé (27)	Maladie (80)
06	Moralité (18)	Immoralité (31)	28	Responsabilité (21)	Irresponsabilité (12)
07	Clarté (21)	Confusion (140)	29	Honnêteté (49)	Malhonnêteté (60)
08	Protection (39)	Nuisance (42)	30	Raison (44)	Folie (88)
09	Confort (24)	Irritation (51)	31	Humilité (25)	Fierté (54)
10	Franchise (33)	Sournoiserie (38)	32	Sécurité (33)	Insécurité (24)
11	Coopération (34)	Conflit (90)	33	Amusement (65)	Horreur (74)
12	Paix (44)	Violence (39)	34	Altruisme (43)	Egoïsme (16)
13	Courage (28)	Lâcheté (48)	35	Innocence (16)	Culpabilité (27)
14	Persuasion (56)	Obligation (45)	36	Sensibilité (34)	Insensibilité (47)
15	Création (38)	Destruction (139)	37	Intelligence (55)	Stupidité (95)
16	Plaisir (77)	Douleur (63)	38	Force (55)	Faiblesse (95)
17	Désir (77)	Evitement (46)	39	Justice (22)	Injustice (37)
18	Louanges (49)	Injures (88)	40	Succès (29)	Echec (39)
19	Energie (57)	Fatigue (100)	41	Vie (36)	Mort (77)
20	Prévisibilité (29)	Surprise (69)	42	Supériorité (35)	Infériorité (44)
21	Excitation (178)	Ennui (102)	43	Abondance (68)	Manque (36)
22	Promesse (34)	Avertissement (54)	44	Admiration (75)	Dénigrement (90)

Table 1: The 44 affect axes chosen for our classification experiment. Each axis has a positive and a negative pole. In parentheses is the number of words manually affected to each axis pole.

Choose Seed Words

1. Avantage (advantage)

- ✓ Avantage
- ✓ Avantageux
- ✓ Avantager

2. Désavantage (disadvantage)

- ✓ Désavantage
- ✓ Désavantager
- ✓ Désavantagée
- ✓ Défavoriser
- ✓ Défavorisée

- Find prototypical noun, adjective, verb
- Expanded using synonym dictionary and manual filtering



1. SL-PMI : Semantic Likelihood Pointwise Mutual Information from Information Retrieval

- ✓ Using the SemanticMap, a resource built from the Web.

$$\text{SL-PMI}_C(w) = \frac{1}{|C|} \sum_{c \in C} \log_2 \frac{\epsilon + H_\delta(w, c)^2}{\epsilon + H_\delta(w, *) H_\delta(c, *)}$$

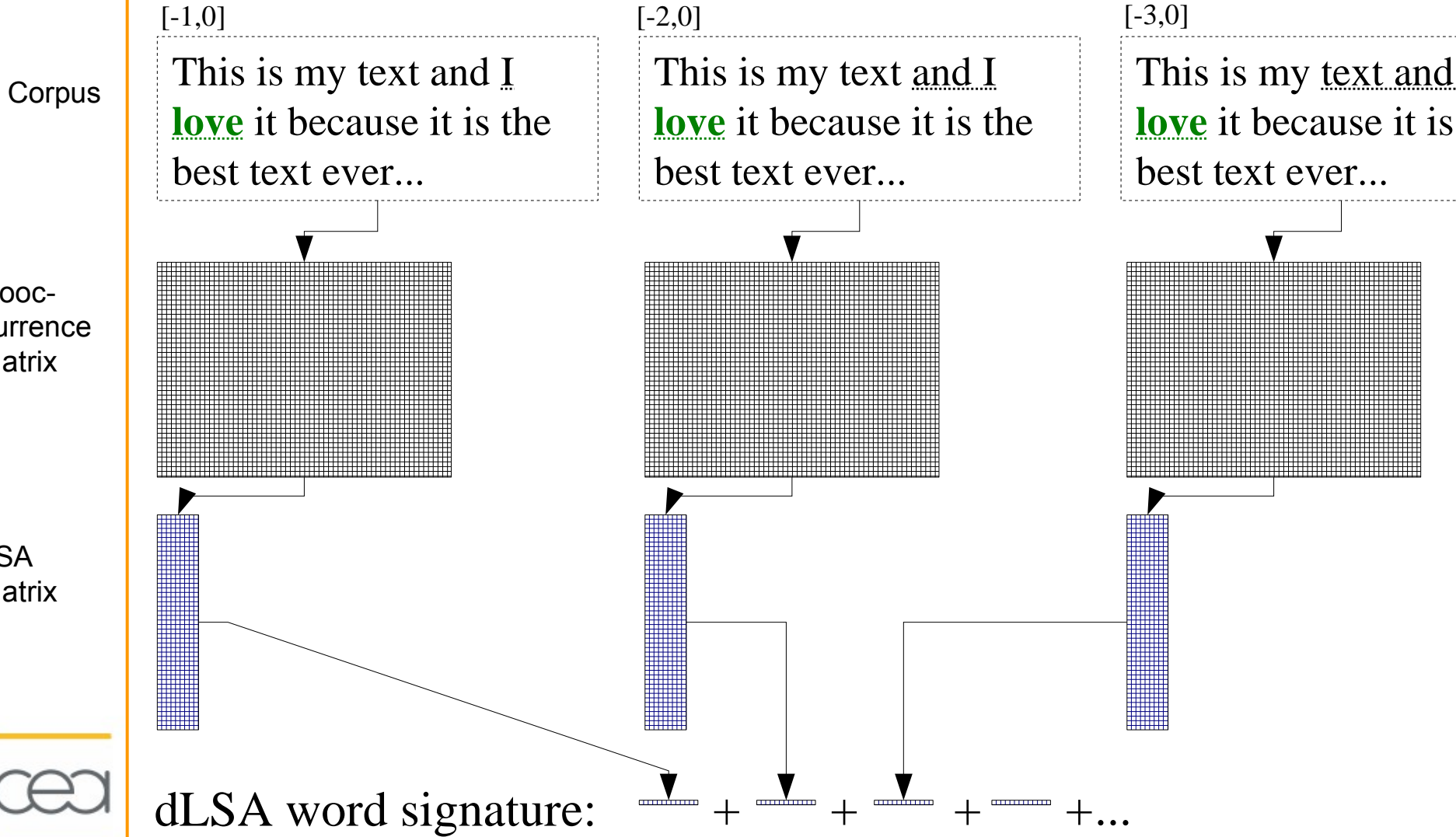
Where $H_\delta(w_1, w_2)$ is the number of cooccurrences of words w_1 and w_2 in a δ words window.

2. SL-LSA : Semantic Likelihood using LSA similarity measure

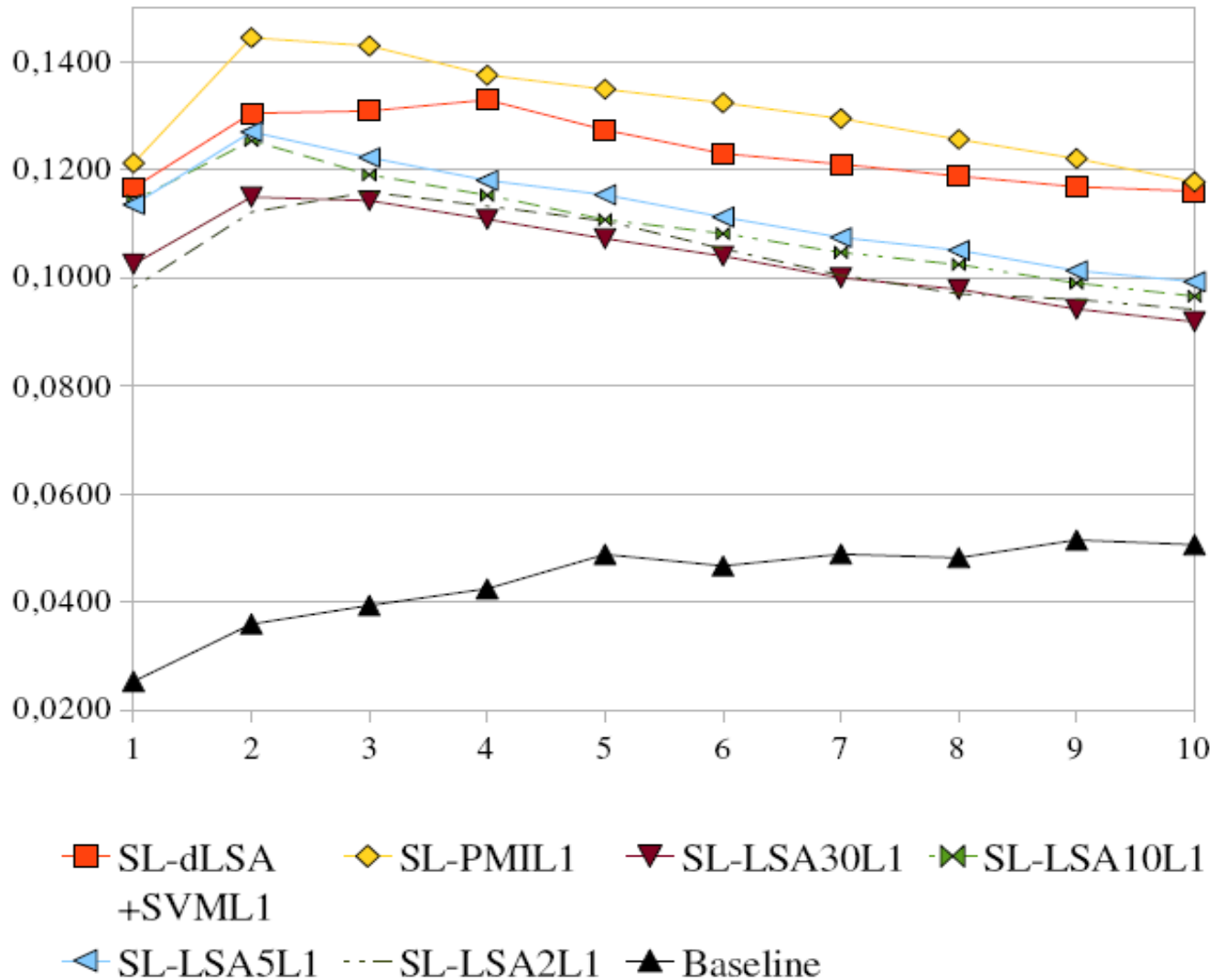
- ✓ Average cosine distance
- ✓ With windows : [-2,+2], [-5,+5], [-10,+10], [-30,+30]
- ✓ Using InfomapNLP + Europarl/French

3. SL-dLSA+SVM : Semantic Likelihood from diversified Latent Semantic Analysis (LSA) and Support Vector Machines (SVM)
 - ✓ Create forty-two 300-dimension LSA spaces
 - **Varying window size (14) × symmetry (3)**
 - Window size: $\delta = [1 \dots 10, 15, 20, 25, 30]$
 - Windows : $[0, +\delta]$ $[-\delta, +\delta]$ $[-\delta, 0]$
 - ✓ Concatenate spaces for each word (12600 dim)
 - ✓ Train a 44-classes SVM classifier

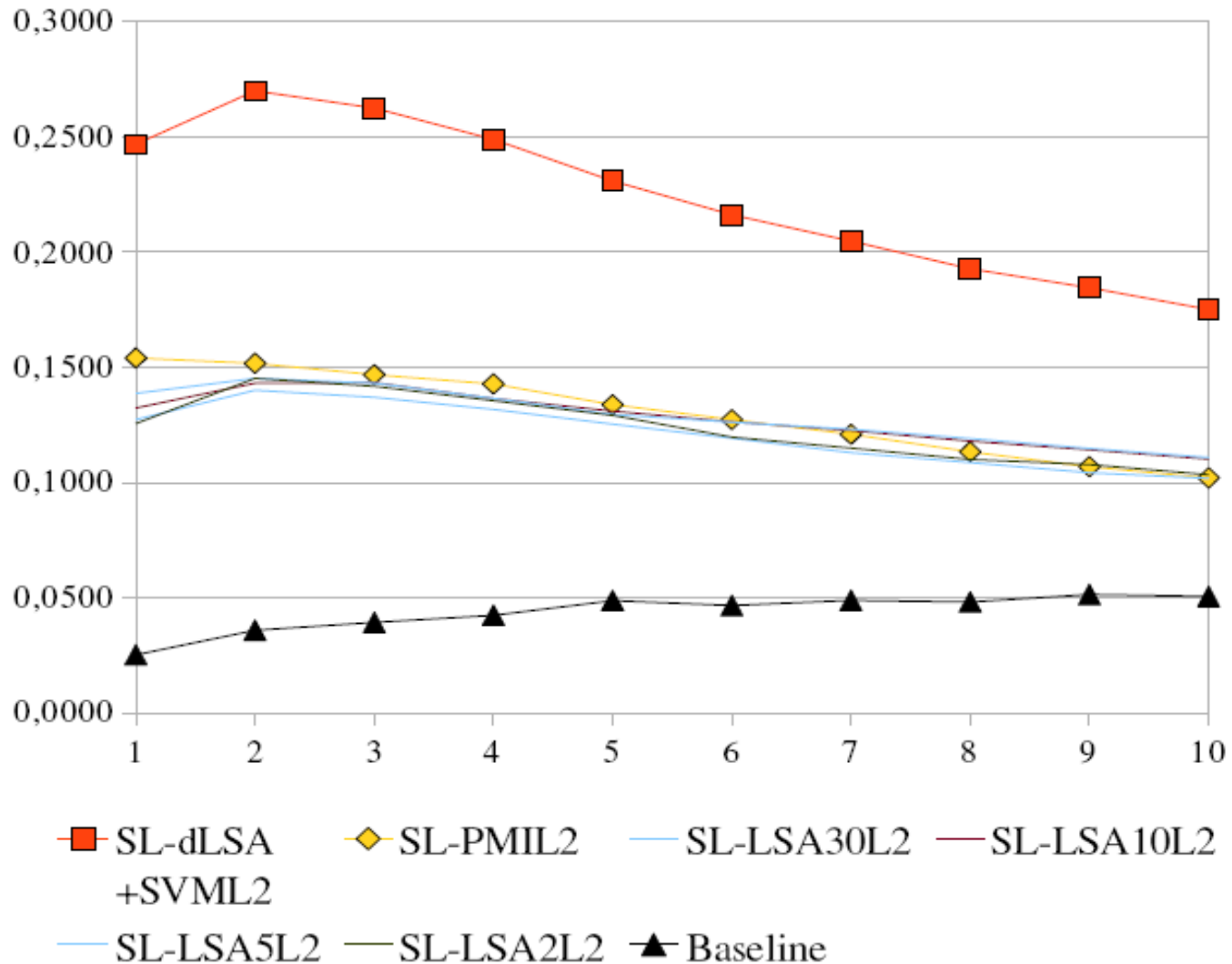
SL-dLSA+SVM



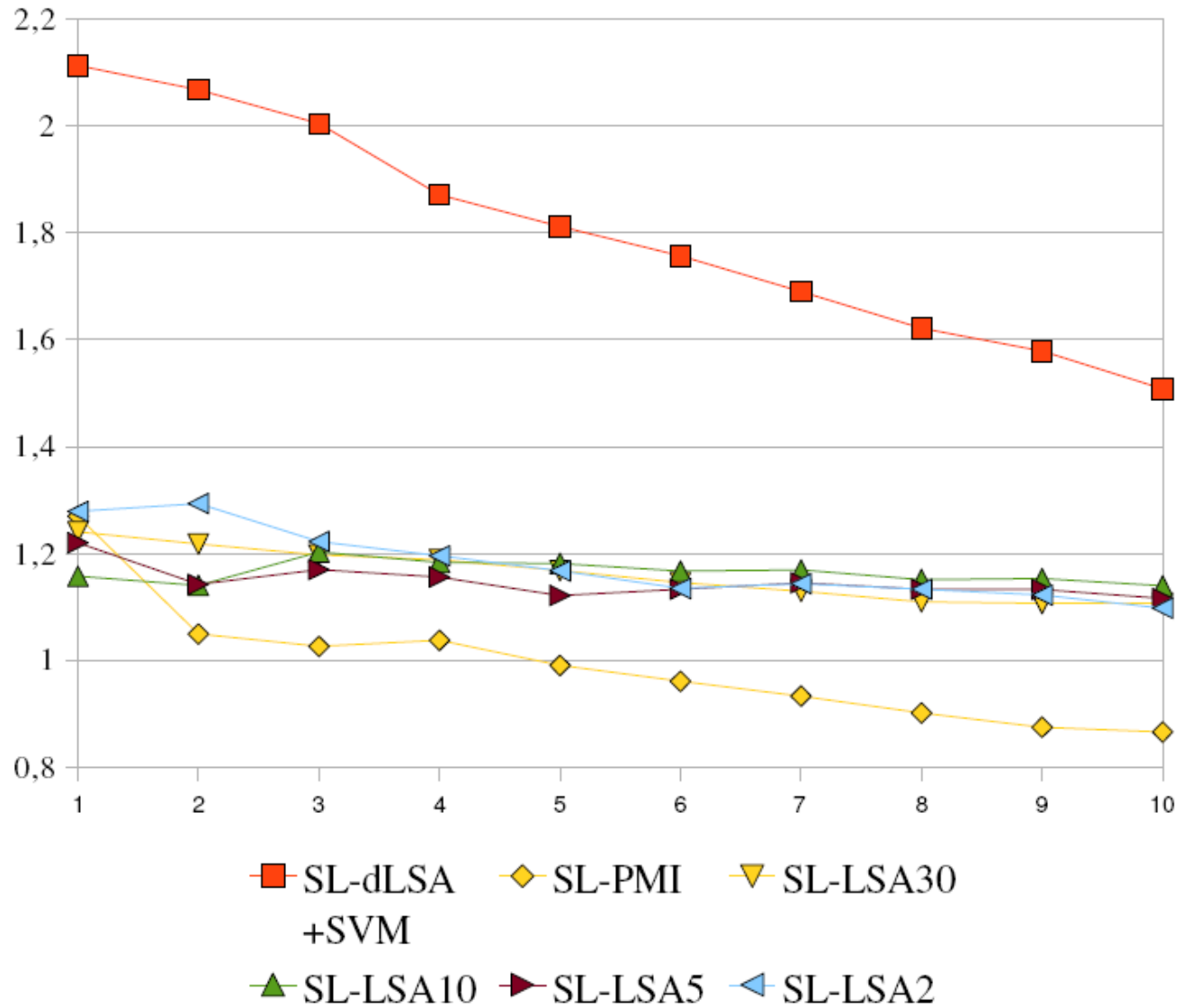
Evaluation: Using five seed words per class



Evaluation: Using twenty seed words per class



Improvement (from 5 seeds to 20)



Good Example of Classifying a New Emotive Word

<i>Class</i>	<i>Score</i>
<i>27 Health/Sickness</i>	.105
01 Advantage/Disadvantage	.065
<i>09 Comfort/Irritation</i>	.065
<i>07 Clarity/Confusion</i>	.062
<i>22 Promise/Warning</i>	.056
<i>36 Sensitivity/Insensitivity</i>	.042
<i>03 Amity/Anger</i>	.040
16 Pleasure/Pain	.037
<i>21 Excitement/Boredom</i>	.035
<i>24 Public-spiritedness/Crime</i>	.034

Table 4: Classification of the word “**désagrément**” using SL-dLSA+SVM with L2, $\#Classes = 10$. Based on the gold standard L3, gold standard classes (in bold) for this word were 01 (Advantage/Disadvantage) and 16 (Pleasure/Pain). Other a posteriori acceptable classes numbers are in italic.



Negative Example

<i>Class</i>	<i>Score</i>
<i>44 Admiration/Denigration</i>	.144
<i>18 Praise/Slander</i>	.101
37 Intelligence/Stupidity	.082
13 Courage/Fear	.071
21 Excitement/Boredom	.062
16 Pleasure/Pain	.035
24 Public-spiritedness/Crime	.029
06 Morality/Immorality	.028
33 Humor/Horror	.028
23 Facilitation/Prevention	.021

Table 5: Classification of the word “**disgrâce**” using SL-dLSA+SVM with L2, $\#Classes = 10$. Based on the gold standard L3, the manually assigned class for this word (01 Advantage/Disadvantage) is absent. A posteriori acceptable classes numbers are in italic.

Conclusions

- 1. An affect dictionary can be built rapidly for a new language using a little manual labor and semi-automatic techniques over a large corpus**
 - ✓ **Best method : 10 times better than baseline**
 - ✓ **Learning from 20 words per semantic axis is better than 5 (for all methods)**
- 2. Semantic Likelihood (SL) from diversified Latent Semantic Analysis (dLSA) and Support Vector Machines (SVM) benefits more from more learning data than SL-PMI or SL-LSA**
 - ✓ **Because of SVM vs. other methods ?**
 - ✓ **Because of the many concatenated LSA spaces ?**



The End



Perspectives

1. Though overall precision rates are comparable, different windows sizes for SL-LSA select different types of similarity, e.g.

- Small windows : synonymous adverbs
- Large windows : same domains

✓ **Explains the results of SL-dLSA+SVM**

2. Questions

- ✓ **Can different window sizes be combined for other problems (disambiguation, alignment)**
- ✓ **Can we combine various SL-LSA ?**

