Multimodal Conformity of Expression between Blog Names and Content

Carl Vogel
Centre for Computing and Language Studies, Computational Linguistics Group
School of Computer Science and Statistics
Trinity College Dublin, Dublin 2, Ireland
email: vogel@tcd.ie

***Abstract*—Using a sentiment lexicon for German translated into English, Italian, Spanish and Swedish, Tumblr URL names for blog archives are inspected with regard to the category of the content. Across the five languages, where the content would be of concern to anyone who would be inclined to filter offensive content, the URLs are constructed with negative sentiment expressions more often than positive sentiment expressions. Other relations suggestive of either conformity or opposition in expression between the modalities of blog name and blog content in image in text are not strongly visible in the 1000 URL sample studied.**

***Index Terms*—multimodal conformity of expression, offensive content, content filtering, blogs, Tumblr**

I. Introduction

The literature on consistency of message across modalities of communication often addresses propositional content [1]. Others also find interesting the extent to which affective content that accompanies propositional content is conveyed across modalities [2]. While there is considerable discussion of what cross-modal relations are relevant to study [3], [4], basic relations of conformity, contradiction and association appear most relevant, whether the content is propositional or affective.

One question to be addressed is whether cross-linguistically, the sentiment associated with terms is manifest in cross-modal content that users index by those terms. In constructing a blog identifier in Tumblr or related services, a blogger must make a choice. Where that choice is a general word or phrase, as opposed to a proper name, it seems safe to assume that the expression is not random, but rather provides some clue to the organization of content developed at the blog. If a blog wears “disgust” in its name, and the content contained in the blog lives up to that description, then there is multimodal conformity. In comparing blog names to content hosted, one is testing conformity of message across modalities that the blogger chooses to use. This work seeks to construct a method for assessing the degree of multimodal conformity in relation to alternatives such as ironic divergence or random patterns.

Apart from the general exercise of classifying human behavior in these terms, a follow-on benefit may obtain. If there is a strong positive correlation between negative sentiment expressions and the “negative” content hosted at websites with names derived from those expressions then there is an easy filter that could be applied to suppress mention of those URLs in online fora where moderation is desired. More generally, filtering is available if properties of the site name in itself give reliable clues about the content. It is easy to imagine situations in which such suppression might be sought, for example if a technology company wishes to filter user-generated content that is located in its user support fora. Such fora have a tendency to develop as online communities, with occasional discussions that wander away from the core topic of the forum. In the course of such discussions, it may well be that users post links to other sites that are friendly (as well as links to sites that relate to the technology content under discussion). The host of the site may choose to ban all such links, or the site may wish to be less proscriptive, implementing intelligent filters. The work reported here is one step in the evaluation of the viability of such filters.

II. Multimodal Matching

A. Method

A German sentiment lexicon was obtained [5]. The lexicon grades sentiment between -1 and 1. The original list contains over 1818 base form negative words and 1650 with positive sentiment scores; with synonyms, the total lexicon exceeds 31,000 items. From each base form list, 100 items were selected at random, without replacement. Tumblr archives were tested on the basis of these 200 expressions. For any given *item*, the corresponding URL, item.tumblr.com/archive, was manually inspected and classified. Multiple categories were used, but any one blog received a single label. The hypothesis was that negative terms would lead to more offensive web archives, and that positive terms would lead to content archives unlikely to offend. Assessing content in relation to the sentiment-scored words implements a kind of multimodal message compatibility check.

Using an online translation service, the 100 terms were translated into other languages: English, Spanish, Italian and Swedish. This was in order to address at least the languages covered by the cross-linguistic emoticon studies of [6], [7]. The idea is that the translated word is likely to have a comparable sentiment in the translation target language. By looking at the content associated with translations of a word with a certain sentiment score, one can consider the multimodal compatibility check across language groups. Of course, it is possible that the translated words, if assessed within language using the methods of [5] would not have sentiment scores

978-1-4799-1546-0/13/$31.00 ©2013 IEEE

identical to the German scores. On the other hand, there is no prior reason to anticipate radical differences in scores, except for considerations raised below. Tumblr archives were surveyed on these original and translated expressions.

A potential confound in the data may arise from that translations could yield "false friends" in one language or another. Additionally, some words lacked direct translation from the original material, and the default technique of the online translation service used appears to be to use English as a pivot language where more informed direct translations are not available. In some cases, a back-off strategy of the online MT service results in English words being proposed as translations in other languages. Where this occurred, the English words were inspected, and tested further to verify that they were not coincidentally "true friends". Where this did not appear to be the case, synonyms were sought. Orthographically, no words with umlauts or other diacritics appear to form Tumblr titles. Where they appeared in words, the resulting pages would not be found. In these cases, systematic alternatives were sought (e.g., *fatastico* in place of *fatástico*;[1] in the cause of umlauts, first the page without the umlaut was sought, and then the page with a long form — "ae", then "a", instead of "ä", etc.).

Fig. 1 provides a baseline for interpreting illustrations that follow. The scale on the right of the level plot provides a coding of sentiment levels witnessed in the sample based on shading, with darker levels indicating magnitude of negative sentiment [8]. The y-axis is labelled according to the 200 item numbers. Thus, the plot illustrates what is stated above, namely, that the word or phrase used in each language is accorded the same sentiment as the German counterpart.

If pages examined were non-empty, distinctive images were probed to form coarse-grained classification of the audience on the basis of user-identification labels for those who left comments. This allows an assessment of the nature of the content and its audience appeal. An iterative process of checking comments on pages with content and then the pages of the commenters is applied. However, only a few pieces of information were recorded for each URL: the coarse grained classification, an assessment of the quantity of postings, and the year of last posting to the account. The categorization was iterative, in that after an initial labelling of the content of each of the 1000 URLS, the profiling of the dataset according to those categories yielded many singleton categories (one that applied to the images at a single URL, however many images were there). The category *sex* was used if images included graphic depictions of the act (or in one case, textual description of prostitution in Kiev) even if this did not describe the largest quantity of items in the blog.[2] One URL specifically about poker was classified under sports.[3] In general, classification was subjective, including the generation of classification labels, on the basis of a dominant impression formed from

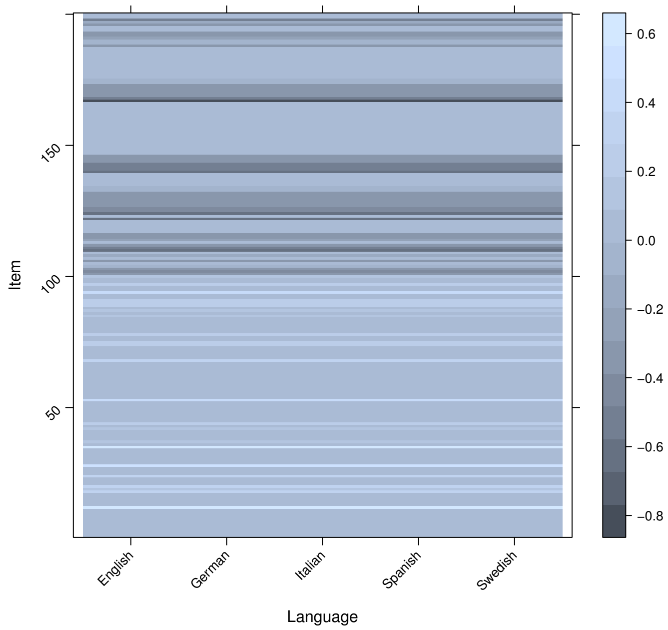

Fig. 1. Constant sentiment values across translation equivalences

images and texts located at a URL. Iterative reclassification proceeded until all singleton categories were either merged with each other or with larger categories (*vintage* (1) with *age* (1), *bikini* (1) with *fashion* (13), etc.). The resulting categories and the number of URLs that fall under each are indicated in TABLE II, along with a recategorization into a smaller set of supercategories.[4] The category *concern* is comprised of those subcategories for which one might anticipate a desire to have successful filtering, in applications such as discussed at the outset. The number of URLs must be interpreted as an estimate: for any URL inspected, if it contained fewer than 100 posts, the figure was counted exactly, and otherwise it was estimated with a finite number of round unit labels (100, 200, 500 or 1000, as appropriate), and therefore a URL may be recorded as having 1000 posts when actually it has ten thousand.[5] Therefore, the counts were treated as an ordered nominal variable, as profiled in TABLE I.

TABLE I
COUNTS OF URLS IN CONTENT-QUANTITY CATEGORIES

(-1,0]	(0,10]	(10,50]	(50,99]	(99,200]	(200,500]	(500,Inf]
569	173	70	22	50	19	97

The count band that includes exactly zero requires further explanation, since it represents the count of URLs with no content. The negative wordlist contained 182 items and the

[1] An exception to this is when the resulting modification evidently creates a word from another language, such as for Swedish *öka* and *ökat*, or German *nervös*.

[2] damper.tumblr.com/archive – last verified: August 2013

[3] raise.tumblr.com/archive – last verified: August 2013

[4] On inspection of the category names alone, one might wonder if *age* is better located under *nature* than *culture*; however, the *age* category itself arose from one URL with many photographs of healthy aged people and another which depicted many objects and people in a way that first suggested *vintage*, and because of this URL the combined category seems to fit better under *culture* than *nature*.

[5] It must be remembered that each of the URLs was inspected manually.

positive wordlist, 198 items, that gave rise to URLs that did not exist (a subset of these required further probing with alternatives for accented characters as described above, but lacked URLs under any of those probes; those that did exist under unaccented vowels were counted in appropriate categories). Further, 86 of the negative and 92 of the positive wordlist URLs existed but contained no content. Finally, eight of the negative and three of the positive wordlist entries pointed to password protected URLs — these were counted as displaying no content. While overall, the number of items in the zero count-band is larger than the number of URLs with content, distribution between negative and positive wordlist items is not significant ($\chi^2 = 1.0439, df = 1$).

The result of classifying URLs in this way is depicted in Fig. 2, which is to be contrasted with Fig. 1. The rows are thicker and fewer in Fig. 2 because URLs derived from individual words of Fig. 1 are grouped under categories here. The sentiment scores associated with the German language words that figure into URLs are constant, but the grouping of these into categories in each of the five languages depends on the categories that seemed suitable for the content supplied by those URLs. Not every category has instances for each language; these are the white areas of Fig. 2. Consider the example of *sex*. None of the URLs for German contained explicit images of sexual acts. Words in Italian that gave rise to URLs containing such content were translations of German words that had relatively strong negative sentiment scores, while English words that led to sexual content were translations of German words with modestly positive sentiment scores. Fig. 2 depicts an additional reclassification, one based on the smaller set of supercategories from TABLE II.[6]

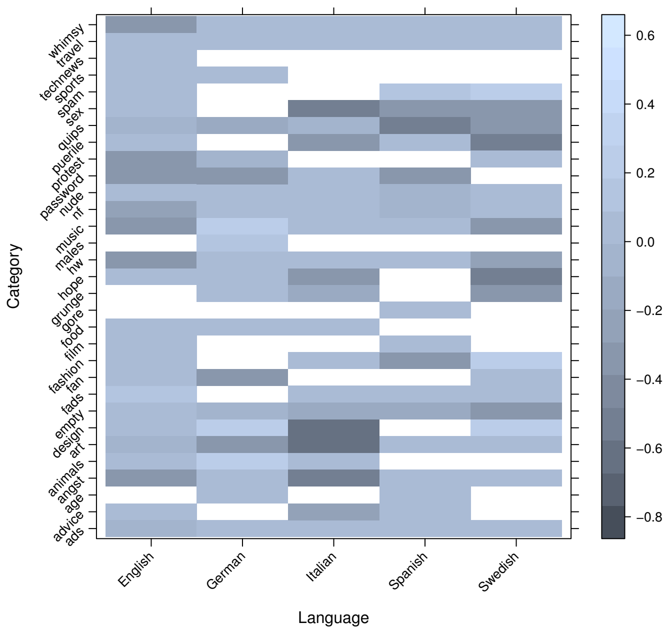

Fig. 2. URL content categories vs. sentiment of translation equivalences

[6]The category *nf* abbreviates *not-found*; *hw* abbreviates *hello-world* and *absent* is synonymous with *not-found*.

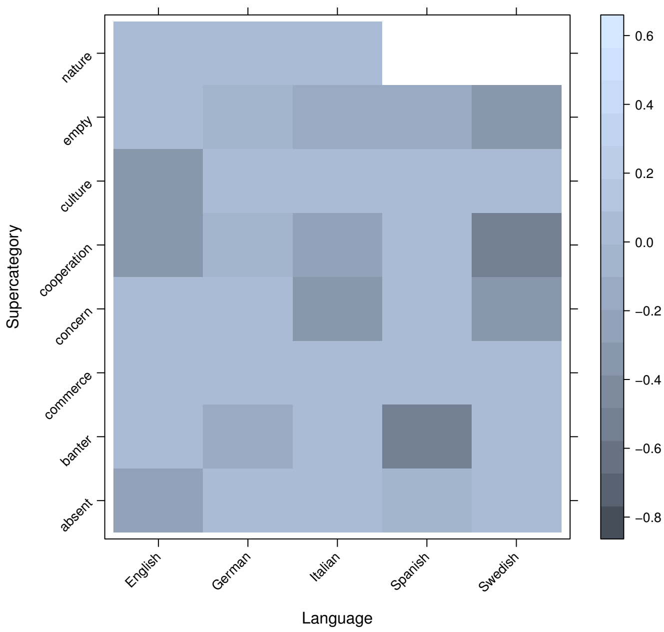

Fig. 3. URL content super-categories vs. sentiment of translations

The data is organized in this way in order to pose questions against it. For example, it has been estimated that 22% of direct links to Tumblr sites were from adult content providers [9]. One might therefore wonder what percentage of Tumblr content depicts sexual activity (explicitly or nearly so). Using the dataset constructed, one may make an approximate response, since the selection of URLs is anchored in sentiment-scored items, rather than random selections of Tumblr URLs without regard to sentiment items in their names. More generally, it is desirable to know whether one can predict, on the basis of sentiment scores for an expression (and language), the category label that will be fit the content of a Tumblr blog which has the expression in its name. This serves as a measure of cross-modal conformity in natural self-expression.

B. Results

1) Quantifying adult content: In many cases, in an attempt to assess the content of a blog on the basis of the blogs of commenters, it appears to be possible to apply the iterative process of checking comments on Tumblr pages and then pages of the commenters with breadth and depth of five (five commenters' pages and five levels from there) before landing on content that one would imagine offensive, disgusting or otherwise disturbing to many [10]. In assessing the proportional content explicitly depicting sexual activity, one can refer to TABLE II to see that 17 of the URLs (1.7%) contained such material. Using the supercategory containing that material, 9.9% of the URLs contained material that one could imagine being filtered by anyone inclined to apply content filters. However, 380 of the 1000 constructed URLs did not exist as Tumblr sites. Thus, 2.74% of the valid URLs contained content depicting explicit sexual activity, and 15.97% contained material categorized as giving *concern*. Thinking of the content not in terms of the number of URLs but the amount of content located at a URL

TABLE II
CATEGORIES AND COUNTS OF URLS BY CATEGORY, AND SENTIMENT POLARITY BY LANGUAGE

Category	URLs	Posting Count	Supercategory	English		German		Italian		Spanish		Swedish	
				-	+	-	+	-	+	-	+	-	+
ads	16	1345	commerce	2	4	0	1	1	3	2	1	1	1
advice	10	1584	cooperation	0	3	0	0	2	0	0	5	0	0
age	2	1180	culture	0	0	0	1	0	0	1	0	0	0
angst	35	7751	concern	6	0	4	3	7	4	10	0	1	0
animals	3	2001	nature	0	1	0	1	0	1	0	0	0	0
art	21	1582	culture	1	2	5	0	1	1	4	1	5	1
design	4	1548	culture	1	0	0	1	1	0	0	0	0	1
empty	178	0	empty	26	4	15	8	10	17	27	25	8	17
fads	17	7928	banter	0	4	0	0	3	3	1	4	1	1
fan	3	52	culture	1	0	1	0	0	0	0	0	0	1
fashion	14	2481	culture	4	5	0	0	0	2	1	1	0	1
film	4	176	culture	0	2	0	0	0	0	1	1	0	0
food	5	96	nature	1	0	1	2	0	1	0	0	0	0
gore	2	1500	concern	0	0	0	0	0	0	2	0	0	0
grunge	5	2214	concern	0	0	0	1	1	1	0	0	1	1
hope	7	99	cooperation	1	0	1	3	1	0	0	0	1	0
hello-world	56	113	cooperation	14	6	3	3	5	8	4	0	8	5
males	2	2000	banter	0	0	0	2	0	0	0	0	0	0
music	15	4832	culture	3	4	0	1	0	1	2	0	2	2
not-found	380	0	absent	6	16	52	53	45	37	21	33	58	59
nude	33	21357	banter	2	3	2	4	5	5	7	4	1	0
password	11	0	password	3	0	1	0	3	1	1	2	0	0
protest	5	214	cooperation	1	0	3	0	0	0	0	0	1	0
puerile	27	13221	concern	5	5	0	0	2	3	4	5	3	0
quips	61	21828	banter	7	14	7	13	2	2	4	8	1	3
sex	17	11852	concern	1	1	0	0	2	1	5	4	2	1
spam	3	181	concern	0	1	0	0	0	0	0	1	0	1
sports	6	115	banter	2	2	1	1	0	0	0	0	0	0
technews	3	794	commerce	3	0	0	0	0	0	0	0	0	0
travel	21	2521	culture	2	2	3	1	3	2	0	5	1	2
whimsy	34	7497	banter	8	0	1	1	6	7	3	0	5	3

(for those URLs that exist, i.e. not falling under *not-found*), TABLE III refers to the item count categories used to classify URLs in relation to the amount of content they contain; of those that have more than 500 items posted, 11.34% contain items depicting sexual activity explicitly (among other items, and in varying proportions among those other items).

TABLE III
PERCENTAGE CONTENT COUNT CATEGORIES SPLIT BY TOPIC

Content	Category	
Item Count	% sex	% other
(-1,0]	0	100
(0,10]	0.58	99.42
(10,50]	0	100
(50,99]	0	100
(99,200]	10	90
(200,500]	0	100
(500,Inf]	11.34	88.66

2) Volume of posts: To help profile the sentiment distribution, Fig. 4 provides violin plots of sentiment assigned to the categorical representation of counts of items posted to URLs for each category. The greater mass of URLs that contain content in the category *sex* is indexed by names that are accorded negative sentiment. Fig. 5 profiles supercategories. These show greater mass of negative sentiment as a function of volume of posting for populous categories.

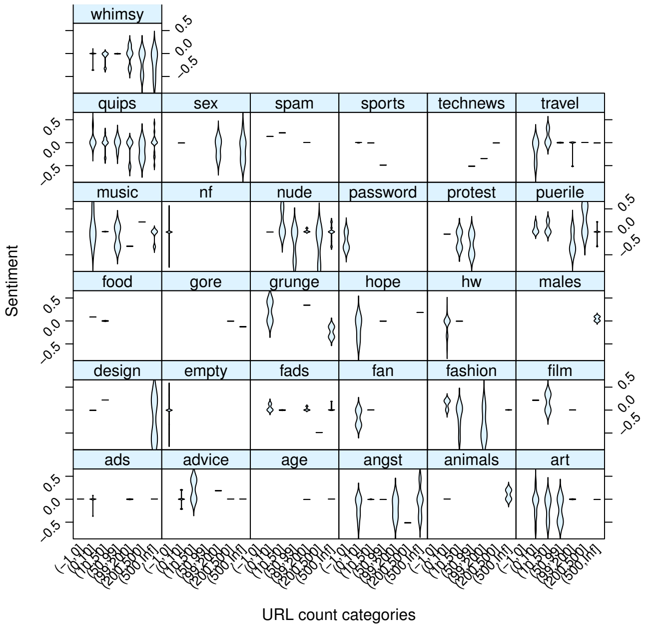

Fig. 4. Count-categories of items posted, by URL sentiment and category

3) Association between category labels and sentiment: If there is relative conformity between the expressions used to name Tumblr blogs and the content of those blogs, then one might anticipate a relation between sentiment scores and the

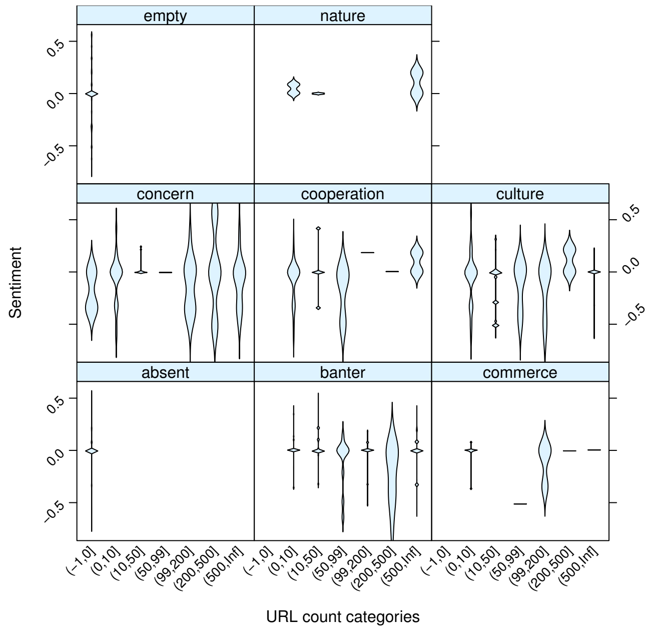

Fig. 5. Count-categories of items posted, by sentiment and super-category

categories used here to label content. It might be that the quantity of posting correlates with strength of sentiment. In order to test the association, this section explores the extent to which the category and supercategory labels used here can be reliably predicted. This is assessed with logistic regression [11], and decision trees [12], [13]. The task here is not to build an optimal classifier based on all known constraints in relation to the data (such as, for example, in the content filtering task that false negatives on some category labels are more costly than others). Rather, the task is to assess whether strong associations between sentiment and language choice on one hand and category labels on the other hand exist.

A decision tree classifier which is considers only the raw the sentiment score and the language, achieves (ignoring the category label *nf*), using 10-fold cross validation, correctly classifies 28.7% of instances using the large set of categories. This is equivalent to the majority classification guesser without further constraints. Using supercategories and not the initial category labels, 27.1% of instances are classified correctly. If the task includes also guessing the non-existence of URLs, then the base category set majority classifier correctly labels 38% of instances (*nf*), while an decision tree labels 39.1% correctly; with supercategories, a decision tree correctly labels 37.3%, again with less than the 38% correct baseline labelling.

As visible in TABLE II, the dataset is such that many combinations of categories are not witnessed at all, and many have a small number of instances, so that systematic differences are not available (of URLs that contain no content in English and German, these tend to be named by expressions with negative sentiment, and in Italian and Swedish, with positive sentiment expressions, but URLs with only "hello-world" postings are named with more negative sentiment expressions in Swedish, Spanish and English and more positive sentiment expressions in Italian; thus, except for the case of English, these trends are opposed but in relation to seemingly very similar overtures). Considering the data through supercategories reveals more structure (TABLE IV): across the five languages, more negative than positive expressions name URLs that are likely to be of concern to anyone who wishing to filter web content.

TABLE IV
CATEGORIES OF URLS AND SENTIMENT POLARITY, BY LANGUAGE

Super-	English		German		Italian		Spanish		Swedish	
category	-	+	-	+	-	+	-	+	-	+
absent	6	16	52	53	45	37	21	33	58	59
banter	19	23	11	21	16	17	15	16	8	7
comm.	5	4	0	1	1	3	2	1	1	1
concern	15	7	5	4	15	10	22	12	7	3
coop.	16	9	7	6	8	8	4	5	10	5
culture	12	15	9	4	5	6	9	8	8	8
empty	26	25	15	8	10	17	27	25	8	17
nature	1	1	1	3	0	2	0	0	0	0

To use logistic regression, a binary variable, CONCERN, is constructed set equal to one if the supercategory of an item is *concern* and zero otherwise. A generalized linear model (with binomial error family) is built to test the influence of language, sentiment and sentiment intensity on CONCERN. The value of sentiment is determined by the original German language expression giving rise to the URLs tested having appeared in the negative wordlist or the positive wordlists. Intensity is the absolute value of the sentiment score recorded in the source sentiment lexicon [5]. Higher order effects were not significant, nor was the effect of sentiment intensity. However, sentiment polarity is significant with negative sentiment items having a greater chance of participating in URLs with content falling in the category *concern* ($\chi^2 < 0.02$) and with a significant influence of language ($\chi^2 < 3.2{*}10^{-5}$), with Swedish and German less likely to have content of concern than English, Spanish or Italian. Within languages, the proportion of URLs that have content of concern versus those that do not is greater for URLs derived from negative wordlist items than for positive wordlist items, but the contrast is not significant within any one language.[7] Content that is fittingly categorized as *banter* seems positive, and in relation to TABLE IV appears, except for Swedish, to arise from more URLs from the positive wordlist than from the negative wordlist. However, this general trend is not significant. For other categories, polarity-related trends are less marked than for *concern* on the negative side and *banter* on the positive side.

III. DISCUSSION AND CONCLUSIONS

The research reported here must be regarded as a first step in analyzing multimodal conformity of expression between blog names and content. Although content at 1000 URLS was examined (by the author alone) after translating 200 German terms into four other languages, this represents 5.8% of the base form German source list. The task is amenable to scaling both with additional human judges, sentiment lexica and social

[7] Adjustments for multiple comparisons are made, conducting the test within the R multcomp package [14].

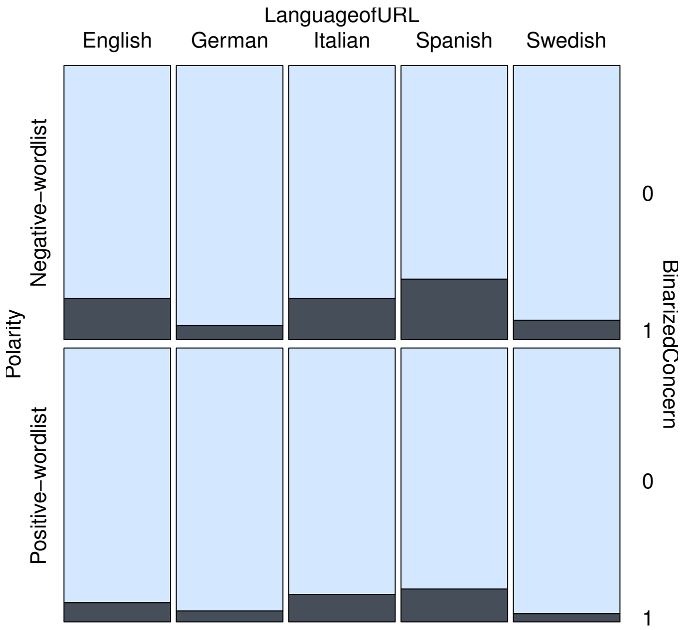

Fig. 6. Proportions of URLs with content of concern (BINARIZEDCONCERN = 1), by language, for URLs constructed from positive and negative wordlists

media providers, as well as incorporation of automated image analysis. The technical means of assessing images that are "streamed" rather than located at a URL, one may presume can be made available. The current work presents a methodology for such scaled experimentation.

A popular media claim [9] has been further quantified. While 22% of links into Tumblr content may be from sites sponsoring adult content, in the non-random sample considered here,[8] just under 3% of URLs contained adult content and approximately 11% of URLs with more than 500 posted contained adult content. This is consistent with the total percentage of items depicting explicit sexual activity posted reaching 22% or more, but seems unlikely.

A contribution of this research is the dataset, which is available to other researchers.[9] The dataset developed includes the translated forms corresponding to the starting-point sentiment lexicon for German, with derived URLs and classifications according to content categorization, coarse-grained quantification of content and annotation of year of last update.

The research here is relevant to applications like large scale off-line indexing of images located on the web [15], but the purpose here is closer to that in which relations between word meaning and image are used to improve overall document understanding [16]. The main finding here supports that this sort of cross-reference in meaning may apply to affective content as well as propositional content. For some categories, e.g. *concern*, there is evidence that relations hold constant across languages, while for others, e.g. *banter*, the choice of language matters more. The direct consequence is in-principle support for selective offensive content filtering on the basis of URL names, without inspection of images.

[8]The sample is constructed from random samples of words that were scored as having either positive or negative sentiment.

[9]See http://www.scss.tcd.ie/Carl.Vogel/MultiModalConformity/.

ACKNOWLEDGMENT

This research is supported by the Science Foundation Ireland funded Centre for Next Generation Localisation (Grant Award 07/CE/I1142 and 12/CE/I2267; www.cngl.ie). Interaction with members of COST Actions 2102 ("Cross Modal Analysis of Verbal and Nonverbal Communication") and IS1004 ("WEBDATANET") have enhanced this work.

REFERENCES

[1] R. Volpe, L. Chanquoy, and A. Esposito, "The matrix of meaning: Representing meaning in mind; prolegomena to a theoretical mind," in *Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues*, A. Esposito, A. Vinciarelli, K. Vicsi, C. Pelauchaud, and A. Nijholt, Eds. Heidelberg: Springer, 2011, pp. 48–55.

[2] M. T. Riviello, A. Esposito, and K. Vicsi, "A cross-cultural study on the perception of emotions: How Hungarian subjects evaluate American and Italian emotional expressions," in *Cognitive Behavioural Systems*, A. Esposito, A. M. Esposito, A. Vinciarelli, R. Hoffmann, and V. C. Müller, Eds. Springer, LNCS 7403, 2012, pp. 424–448.

[3] K. Pastra, "Beyond multimedia integration: Corpora and annotations for cross-media decision mechanisms," in *Proceedings of the 5th Language Resources and Evaluation Conference (LREC)*, 2006, pp. 499–504.

[4] ——, "Cosmoroe: A cross-media relations framework for modelling multimedia dialectics," *Multimedia Systems Journal*, vol. 14, no. 5, pp. 299–323, 2008.

[5] R. Remus, U. Quasthoff, and G. Heyer, "SentiWS - a publicly available German-language resource for sentiment analysis," in *Proceedings of the 7th International Language Ressources and Evaluation (LREC'10)*, 2010.

[6] J. Janssen and C. Vogel, "Politics makes the Swedish :-) and the Italians :-(," in *Sentiment Analysis: Emotion, Metaphor, Ontology & Terminology*, K. Ahmad, Ed., 2008, pp. 53–61, poster at presented EMOT2008 Workshop at LREC 2008.

[7] C. Vogel and J. Janssen, "Emoticonsciousness," in *Multimodal Signals: Cognitive and Algorithmic Issues*, M. M. Anna Esposito, Amir Hussain and R. Martone, Eds. Springer Berlin / Heidelberg, 2009, pp. 271–287, COST Action 2102 and euCognition International School Vietri sul Mare, Italy, April 21-26, 2008. Revised Selected and Invited Papers.

[8] D. Sarkar, *Lattice: Multivariate Data Visualization with R*. Springer: New York, 2008.

[9] J. Brustein, "If Yahoo buys Tumblr, what will it do with all that porn?" *Businessweek*, 2013, www.businessweek.com/articles/2013-05-17/if-yahoo-buys-tumblr-what-will-it-do-with-all-that-porn – last verified, August 2013.

[10] C. Vogel, "Intended and unintended offence," 2013, to appear, post-proceedings of the *23rd Italian Workshop on Neural Networks (WIRN13)* in the special session, *Social and Emotional Networks for Interactional Exchanges*.

[11] W. Venables and B. Ripley, *Modern Applied Statistics with S*, 4th ed. Springer, 2002.

[12] J. R. Quinlan, *C4.5: Programs for Machine Learning*, ser. Morgan Kaufmann Series in Machine Learning. Morgan Kaufmann, 1993.

[13] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. Amsterdam: Elsevier, 2005.

[14] F. Bretz, T. Hothorn, and P. Westfall, *Multiple Comparisons Using R*. Chapman & Hall/CRC: Boca Raton, Fl., 2011.

[15] X.-J. Wang, Z. Xu, L. Zhang, C. Liu, and Y. Rui, "Towards indexing representative images on the web," in *Proceedings of the 20th ACM international conference on Multimedia*, ser. MM '12. New York, NY, USA: ACM, 2012, pp. 1229–1238.

[16] E. Bruni, J. Uijlings, M. Baroni, and N. Sebe, "Distributional semantics with eyes: Using image analysis to improve computational representations of word meaning," in *Proceedings of Multi-Media 12 (20th ACM International Conference on Multimedia),*. ACM, 2012, pp. 1219–1228.