

On Principles of Annotated Databases of the Semantic Field “Aggression”

Rodmonga Potapova and Liliya Komalova

Moscow State Linguistic University, Institution for Applied and Mathematical Linguistics,
Moscow, Russia

{RKPotapova, GenuinePR}@yandex.ru

Abstract. The article describes main principles of elaboration of annotated databases of the semantic field “aggression” and describes Russian and English Digital Mass Media Full-Text Databases containing verbal representatives of the concept “aggression”©. Each database consists of 120 manually annotated text units. The annotation covers lexical, semantic and pragmatic levels. Special metrics and local dictionary of the semantic field “aggression” accompany each text. The databases can be implemented in scientific research on Speechology, computer-aided Internet monitoring systems teaching, in educative process, for further elaboration of searching systems based on the semantic field “aggression”.

Keywords: Corpus linguistics, verbal aggression, annotated database, semantic field “aggression”, digital mass media.

1 Introduction

Digital mass media texts as usual serve for building public opinion and aim at modeling recipient’s point of view on real facts. Positive and negative emotions on pages stimulate similar emotions and modalities of a recipient. That’s why it’s of great importance to monitor this sector and elaborate special automatic Internet monitoring systems based on different thematic databases of emotionally colored texts.

In the framework of the survey of the semantic field “aggression” we focus on means and mechanisms of motivation and realization of negative emotions and modalities, including those provoked mediate texts of digital Internet mass media. Linguistic analysis of modern publications in European languages (English, Spanish, Russian languages) confirms that nowadays semantic field “aggression” in mass media texts is widening and the perception criticality threshold is reducing. However, accessibility of digital texts and modern programming means allow monitoring Internet textual content and revealing texts provoking aggression, extremism and separatism.

In this context elaboration of special thematic verbal databases and knowledge bases which can be implemented in Internet filter systems becomes perspective way to cluster language and speech texts’ tendencies of modern digital Internet mass media along with the possibility of assessment of their effecting potential to different forms of aggressive behaviour.

2 Method and Procedure

Necessary request for the first phase of database constructing is to provide a scientific research resulting in the list of characteristics describing how concept “aggression” is functioning on pages of digital Internet mass media. The next phase is material selection which must be homogeneous: it must contain topics with approx equal number of text units each, representing particular time period. The list of sources determines attitude of the editors and therefore directives and stereotype concepts translated to the audience.

Once the material is collected it must be correctly described. Description gives understanding of social context in which the corpus was made. It’s significant socio-cultural index, especially for users of other lingua-cultures. It also gives clue to the next phase – annotation.

Our Russian and English Digital Mass Media Full-Text Databases© have identical annotation form and contain verbal representatives of the concept “aggression”. Each of them was gathered from authentic sources: articles from digital analogues of newspapers, scripts of TV-news programmes, messages of news agencies, web-site news kits (time period: 2011-2013). Results of the survey and principles used for marking words related to state of aggression are discussed in [2,3]. Each unit of both databases consists of the text and its annotation:

I. Date-line of the text:

1. Headline,
2. Name of the source,
3. Date of publication,
4. Name of the author,
5. Web-site address;

II. Linguistic passport of the text:

6. Main topic of the text (choose one of the following: geopolitical conflicts, political conflicts, military operations, criminal behaviour, personal violence, family violence, aggressive economics, hearings, extremist acts, racism, inter-confessional dissention, xenophobia, call to violence and discrimination),

7. Genre (choose one of the following: brief article, report, article, leaflet, commentary, satirical article, interview, essay),

8. Pragmatic component (choose one of the following: informing, analysis, polemics, interpretation, announcement, agitation, propaganda, criticism);

III. Each text has special metrics showing:

9. Total amount of notional and non-notional words in the text,
10. Number of graphemes without blanks,

11. Annotation also includes local dictionary of the semantic field “aggression” (a list of words from the text which belongs to the semantic field “aggression”) which consists of corresponding words from a particular text. All these words are marked in the text.

12. Density of words from the semantic field “aggression”. Table 1 and table 2 give examples of annotated text units.

Each database consists of 120 manually annotated text units selected out of 4000 articles sample from 29 Russian and 43 English sources (table 3). To analyze texts we use content-analysis algorithm supplemented by method described in [5,6] and recom-

Table 1. An example of annotated text unit in Russian Digital Mass Media Full-Text Database

Название текста	Президент Сирии ужесточил наказание за похищения и убийства людей
Название источника	Информационный Интернет-портал РИА-Новости
Дата публикации текста	02 апреля 2013 года
Автор текста	Денис Малков
Ссылка на источник	http://ria.ru/arab_sy/20130402/930583218.html
Тема текста	Основной тематической составляющей данного текста является похищение. Об этом свидетельствует абсолютная доминирующая позиция следующих слов и словосочетаний «похищения», «похитил», «похититель», «смерть похищенного», «массовые похищения» в составе семантического поля «агрессия» для данного текста.
Жанр текста	Заметка
Прагматическая составляющая текста	Информирование
Количество слов	151
Количество знаков (графем) без пробелов	969
Слова, относящиеся к семантическому полю «агрессия»	Ужесточение наказания, похищение людей, похитить человека, приговорить, пожизненные каторжные работы, причинения вреда здоровью, изнасилование, смерть, похищенный, похититель, казнить, наказание, сроки заключения, массовые похищения людей, повстанцы, криминальные группировки, похищенный, вооруженный конфликт, оппозиция, погибнуть, противостоять организованным действиям, вооруженный, боевик, наемник
Плотность СПА	25,16%
Исходный текст	<p>Президент Сирии Башар Асад подписал указ об ужесточении наказания за похищение людей, сообщает во вторник пресс-служба главы государства.</p> <p>"Каждый, кто похитил человека, будет приговорен к пожизненным каторжным работам, а в случае причинения вреда здоровью, в случае изнасилования или смерти похищенного, похититель будет казнен — говорится в указе сирийского президента.</p> <p>Ранее наказание в зависимости от последствий предусматривало различные сроки заключения.</p> <p>Массовые похищения людей в Сирии отмечаются на территориях, контролируемых повстанцами, суммы выкупа, которые криминальные группировки требуют от близких и родственников похищенных, разнятся от нескольких тысяч до нескольких миллионов долларов.</p> <p>Вооруженный конфликт в Сирии между властями и оппозицией продолжается с марта 2011 года. По данным ООН, за это время в стране погибли около 70 тысяч человек. Сирийские власти заявляют, что противостоят организованным действиям хорошо вооруженных и обученных местных боевиков и наемников из-за рубежа.</p>

Table 2. An example of annotated text unit in English Digital Mass Media Full-Text Database

Headline	When violence hits a nerve
Name of the source	The Washington Post
Date of publication	06.02.2013
The author of the text	Esther J. Cepeda
Web-site address	http://www.jsonline.com/news/opinion/when-violence-hits-a-nerve-2b8lpp9-190094521.html
Main topic of the text	Critics of people’s attitude toward violence in case of a girl murder
Genre	Critic article
Pragmatic component	Analysis
Total amount of notional and non-notional words	672
A number of graphemes without blanks	3123
Local dictionary of the semantic field “aggression”	Violence, hit a nerve, dangerous, corrupt, racially divided, murder, outrage, unspeakable, violence, victim, death, strike a chord, hit a nerve, to die, woe, victim, rough, die, frighten, clout, power, quell, violence, gun down, tough, murder, homicide, interminable, murder, spree, violence, homicide, victim, troops, kill, Afghanistan, tragedy, post-mortem, victim, gun, violence, kill, gunman, die, gunman, shoot, heartbreak, victim, tragedy, ugly, violence, brushes with the law, gang affiliations, hurt, shoot , die, bad, horror, shooting, rallying cry, prevent horror, violence, rip apart, struggling, ache, disadvantaged, at-risk, victim, senseless, violence, fall short, gun-related, violence, bad, harm, tragic, harm
Density of words from the semantic field “aggression” in the text	11%
Full text	<p>Chicago - To a lifelong resident of one of America’s most dangerous, corrupt and racially divided cities, the murder of Hadiya Pendleton has brought to light what really drives coverage – and outrage – of unspeakable violence: a virtuous victim.</p> <p>Did a 15-year-old’s death really strike a chord nationally because she had, a week before, been to our nation’s capital performing at the inauguration festivities? No, that wasn’t it – nor did her story hit a nerve because of where she died.</p> <p>The typical woe of victims in the rough parts of town is that their stories don’t get reported in as much detail as those who die in a nicer parts – the neighborhoods that, when people are frightened, have the clout to make those in power listen to their demands that something to be done to quell the violence.</p> <p>But Pendleton was gunned down in a tough neighborhood – one about a mile from the Chicago home of Barack Obama, a president who has yet to acknowledge publicly that his hometown has gotten an international reputation for being a “Murder City”. The city recorded 550 homicides between January 2012 and the end of last month, so many that most of them have just become footnotes to Chicago’s interminable murder spree.</p>

Obama's South Side neighbors have pleaded with the president to come back and call for national action on inner-city violence. Yet their pleas have been ignored ever since last summer, when the city reached the milestone of having had more homicide victims through the first six months of the year than the number of U.S. troops killed in Afghanistan in the same time period.

Also notable in the tale of this most recent child tragedy is that the teenager didn't rocket to post-mortem attention because of the perennial complaint that victims of gun violence are only acknowledged when large numbers of them are killed by a single gunman. Pendleton was the only child to die after a gunman shot into a crowd of 12 students.

Now, you could make the argument that the national heartbreak was due to the same reason that seems to sort some victims of tragedy into the media spotlight, while leaving others in anonymity: beauty. As one Chicago newspaper put it: "City's ugly violence now has a pretty face".

But that's only partially right. The rest of it is that Pendleton was as close to being as innocent as you can get.

Media accounts tell of a girl who had no brushes with the law or gang affiliations and was an honor-roll student attending a selective-enrollment college prep school. She was an athlete and a majorette in the school's marching band, and a girl with dreams of a profession after getting in and making it through an elite university.

In short, she was the kind of bright child any one of us would be proud to parent or mentor, and it hurts to think of her getting shot in the back and dying in the middle of the street.

These are not bad impulses. They're quite human, and they're the ones credited with turning the horror of the Newtown, Conn., school shootings into a rallying cry that mustered the political courage to look for measures that would prevent such a horror from ever happening again.

Unfortunately, those instincts won't get us far enough in stemming the violence that rips struggling communities apart every day.

Until our hearts start aching for every disadvantaged-at-risk child who becomes a victim to senseless violence – whether perfectly virtuous or not – our aspirations for safe communities will be sure to fall short.

Far too many of us have become too accepting of the steady stream of gun-related violence that happens in "bad neighborhoods", "somewhere else" or only to "bad people". It's time to start believing that a harm to the least of us is every bit as tragic as a harm to the rest of us.

recommendations given in [4]. Density of words from the semantic field "aggression" was calculated with the open license programme "Textus Pro 1.0" [1].

3 Conclusions

Russian and English Digital Mass Media Full-Text Databases can be implemented as:

- wordforms databases of the semantic field "aggression" for computer-aided Internet monitoring systems teaching to reveal potential provocative/conflictive messages in official digital Internet mass media sources and also to determine places of geopolitical "hot spots" in the world;
- wordforms databases for thematic frequency dictionary elaboration in Russian and English;

Table 3. Russian and English digital Internet-sources

Russian sources	British English sources	American English sources
	Digital versions of newspapers	
aif.ru gazeta.ru izvestia.ru kp.ru mk.ru mn.ru novayagazeta.ru rg.ru tambov.mk.ru utro.ru	belfasttelegraph.co.uk cambridge-news.co.uk express.co.uk dailymail.co.uk edp24.co.uk ellesmereportpioneer.co.uk nwemail.co.uk socialistworker.co.uk icbirmingham.icnetwork.co.uk prestoncitizen.co.uk guardian-series.co.uk guardian.co.uk yorkpress.co.uk voice-online.co.uk theweek.co.uk walesonline.co.uk	chicagotribune.com ctpost.com dailyherald.com democratandchronicle.com freep.com examiner.com lasvegassun.com latimes.com jsonline.com nypost.com newsok.com post-gazette.com politicshome.com theadvocate.com bostonglobe.com economist.com mcall.com nytimes.com thelede.blogs.nytimes.com nytsyn.com observer.theguardian.com seattletimes.com tbo.com washingtonpost.com
Russian sources	British English sources	American English sources
	News agencies	
regnum.ru ria.ru top.rbc.ru itar-tass.com	–	ap.org reuters.com
	Internet-portals	
komionline.ru lenta.ru newsru.com online812.ru rbcdaily.ru ridus.ru news.mail.ru news.rambler.ru memo.ru	–	news4theworld.com 7.politicalbetting.com
	Broadcasting company news-lines	
echo.msk.ru rus.ru ruvr.ru vesti.ru ntv.ru	ru.euronews.com bbc.co.uk	–

- wordforms databases for elaboration of searching systems based on the semantic field “aggression”;
- texts databases for scientific research on Speechology to carry out a morphological and syntactical analysis with aim to reveal the text structure which can provoke a recipient’s state of aggression;
- texts databases in educative process within training disciplines as “Linguacnffictology”, “Linguacriminalistic text expertise”, “Fundamental and Applied Speechology”, “Sociolinguistics”;
- texts databases in complex survey of language and speech characteristics of broadcasting and production of aggressive behaviour through digital Internet mass media texts.

4 Prospects of Investigation

Further investigation can be related with elaboration of inner searching system trained to filter Internet-content.

Acknowledgments. The survey is being carried out with the support of Russian Science Foundation (RSF) in the framework of the project No.14-18-01059 at Moscow State Linguistic University (scientific head of the project – R.K. Potapova).

References

1. Kaplunov, D.A.: Textus Pro 1.0, <http://www.blog-kaplunoff.ru/programmy-dlya-kopirajterov.html>
2. Potapova, R.K., Komalova, L.R.: Lingvokognitivnoe Issledovanie Sostoyaniya, "Agressiya" v Mezhyazykovej i Mezhkul'turnoj Kommunikatsii: Pis'mennyj Tekst. In: Semioticheskaya Geterogenost' Yazykovej Kommunikatsii: Teoriya i Praktika. Chast' II, vol. 15(675), pp. 164–173. MGLU, Moscow (2013)
3. Potapova, R., Komalova, L.: Lingua-Cognitive Survey of the Semantic Field "Aggression" in Multicultural Communication: Typed Text. In: Železný, M., Habernal, I., Ronzhin, A. (eds.) SPECOM 2013. LNCS, vol. 8113, pp. 227–232. Springer, Heidelberg (2013)
4. Potapova, R., Potapov, V.: Auditory and visual recognition of emotional behaviour of foreign language subjects (by native and non-native speakers). In: Železný, M., Habernal, I., Ronzhin, A. (eds.) SPECOM 2013. LNCS, vol. 8113, pp. 62–69. Springer, Heidelberg (2013)
5. Potapova, R.K., Potapov, V.V.: Semanticheskoe Pole "Narkotiki": Diskurs kak Obekt Prikladnoj Lingvistiki. URSS, Moscow (2004)
6. Potapova, R.K., Potapov, V.V.: Yazyk, Rech, Lichnost. Yazyki Slavyanskoj Kultury, Moscow (2006)