Behavioral Profile: Synonyms of 'Disagree

Muhammad Zakaria
Nanyang Technological University

Abstract

There have been some developments in the area of lexical semantics through corpus linguistics in recent time. As a result, it is now possible to focus on semantic dimension of lexemes from a usage-based perspective. This paper presents a Behavioral Profile of synonymous words of the verb disagree. The data for this paper is taken from the Global Web-Based English (GloWbE) corpus.

Keywords

Corpus linguistics, behavioral profile, synonyms, GloWbE.

Introduction

Empirical data bears semantic profiles of lexemes in utterances, which are affected by situational and cultural contexts. In other words, an utterance is constrained by the conventions or codes of a society. There are many approaches which linguists follow to study the lexemes in empirical data in order to find out the relevant semantic analysis of the lexemes. Behavioral Profile (henceforth, BP) is a recent development in corpus linguistics based on corpus-based data, which investigates distributional characteristics of lexemes. The approach of BP can be related to the works of Divjak and Gries (2009). In defining the BP approach, Dobrić (2010: 98) states that,

> ...behavioral profiling presents a combination of corpus methodology as a practical orientation and cognitive and sociosemantic theoretical oriental. The theoretical background of the procedure can be found in the combination of what is basically a (manual) cognitive approach to sense identification (in the initial steps of the procedure) and a sociosemantic, in its more diluted form, conception of meaning (regarding the observation of the context as generating sense recognition).

This paper aims to build the semantic understanding of the synonyms of the verb disagree (for BP of synonyms and antonyms, see Gries and Otani 2010). There have been several studies on the verb disagree, specially its conceptual definition (Koczogh 2013). Koczogh (2013: 213) states, "Disagreement and argument are the most frequently used terms in the area of conflict talk…" Thus, the verb disagree is related to the conflicting and argumentative context. The synonyms of disagree enables the speaker's power to be different in his/her action and opinion. Studies on other synonyms of the verb disagree used in this paper includes Kakavá (2002) on opposition and Kotthoff (1993) on dispute. This paper focuses on the distributional features of these lexemes with a detailed behavioral discussion through semantic and syntactic factors. The methodology of this research paper is described in §2 . In §3, I have presented a BP analysis based on a data frame prepared in the methodological stage (see §2). I have listed down the findings of this BP analysis in §4, which is followed by some conclusive remarks in §5.

Methodology

Following a BP approach, we need to go through several steps to build a dataframe before yielding an interesting result out of the dataframe. Three steps are prerequisite to build a dataframe for running a BP analysis (for more, see Gries, 2010: 226-228). The preparatory phases for a BP study include retrieving a sample of concordance lines from a corpus (§2.1), ID tagging (§2.2) and drawing a co-occurrence table (§2.3).

1. Concordance Lines

The first step in the BP approach entails taking a sample of concordance lines of the synonyms (for this study, the synonymous words are - contradict, disagree, dispute, oppose and reject) from the selected corpus. For this paper, I have randomly retrieved 100 sample concordance lines for each of the synonymous verbs from the Global Web Based English (GloWbE) corpus of 1.9 million words based on 1.8 million webpages of 20 English-speaking countries. I have used the data from webpages of Bangladesh. A sample of concordance lines of the investigated synonyms are given from a-c.

  a. …on the topic of emancipation of women, but let me disagree about Asian women. Have you visited progressive countries in...
  b. ...that the headmen of at least three Penan communities that have opposed logging have lost official recognition from Malaysian…
  c. ...love of God and thus be sanctified. Therefore, he contradicted the dogmatic truth in the Council of Trent above that teaches the…

2. ID Tagging

Following the retrieval of 100 concordance lines for each of the verbs, we need to tag different properties of each concordance lines for respective lemmas. Different choices of

Table 1: ID tagging of the concordance lines

Type of factors	Factors	Levels
Morphological	Verb_Form	Present, Past, Infinitive (Inf), Present Participle (Prespart) and Past Participle (Pastpart)
	Mood	Indicative (IND), Subjunctive (SUB), Interrogative (INTRG), Imperative (IMP), Conditional (COND) and Optative (OPT)
Syntactic	Clause_Type	Main, Subordinate and Relative
Semantic	Sem_Sub	Animate, Inanimate
	Sem_Obj	Animate, Inanimate
Pragmatic	Genre	Media, Blog, Forum, Profsite (professional site) and Govserv (government sites)

morphological, syntactic and semantic ID tags could be taken into considerations. This step is manual. I have used the ID tags in Table 1 for the dataframe. Since the ID tagging is manual, sometimes our own judgment is required for tagging different properties of the lemmas in the concordance lines.

The selected corpus does not provide any genre for the sources of the concordance lines. Therefore, I had to go through the sources of each concordance line and decide the genre. I had to consider the over tagging of the data too; for that, I put different sources into one genre. The genre of Media includes web pages of newspapers, magazines, entertainment groups and political or personal information providers or promoters. The genre of Blog includes all personal and professional blogs, while the genre of Forum represents websites of any group for a cause, social welfare groups and non-government organizations (NGO). All the professional sites which promote any business or service are represented by the genre of Profsite, and lastly, the genre of Govserv represents the webpages of the government. I had to face another problem, when I started tagging based on semantic subject (Sem_Sub) and semantic object (Sem_Obj); the problem was about the animacity of the subject or the object. I have considered the lexemes such as the government, political party, etc. as inanimates. The rational for this choice lies in the fact that this type of lexemes represents the abstract entities (in Bangladesh) which are directed by certain ideals. Therefore, I have tagged them as inanimate subjects or objects, as in the sample sentences a and b.

    a. The opposition party has now found yet another point to disagree on, with the ruling government, and blame them for their…
    b.   ...over GMR 's development of the INI airport, which is opposed by all parties of the coalition government. # President Dr.…

The number of such lexemes is scarce in the data frame used for this paper. Therefore, it is fair to claim that the result is not misleading or biased. Table 2 represents a sample of tagged verbs.

Table 2: A sample table of tagged verbs

Verb	Genre	Clause_Type	Verb_Form	Mood	Sem_Sub	Sem_Obj
rejected	Media	Main	Past	IND	Animate	Inanimate
rejected	Forum	Main	Past	IND	Animate	Inanimate
rejected	Blog	Main	Past	IND	Animate	Animate
reject	Blog	Main	INF	IND	Animate	Inanimate
rejects	Media	Subordinate	Present	IND	Animate	Inanimate
rejecting	Forum	Main	PRESPART	IND	Animate	Inanimate
rejected	Forum	Main	Past	IND	Animate	Inanimate
reject	Media	Main	Present	IND	Animate	Inanimate
reject	Media	Subordinate	Present	SUBJ	Animate	Inanimate
rejected	Blog	Main	Past	IND	Inanimate	Inanimate

3. Co-occurrence table

The next step of the BP approach is to convert the data frame into a co-occurrence table. This co-occurrence table shows the number of occurrences for each verb for the levels with which they are tagged in the data frame.

Table 3: A selection of the co-occurrence table for the Genre factor

Verb			Levels of Genre		
	Genre_Blog	Genre_Forum	Genre_Govserv	Genre_Media	Genre_Prof
contradict	27	9	2	50	12
disagree	22	24	1	48	5
dispute	23	12	2	38	25
oppose	17	14	1	63	5
reject	19	20	4	43	14

The co-occurrence table in Table 3 shows the number of occurrences of each verb for each of the levels under each of the factors. Table 3 shows that the verb contradict has 27% occurrence in the genre of Blog, and the maximum percentage of use the verb contradict is in the genre of Media, which is 50%.

Data frame analysis

This section of the paper focuses on the analysis of our data frame based on different packages run in Rstudio. The packages used for this data frame provide a consistent analysis, which tells an interesting story about the synonymous lexemes I have investigated. There are two types of analysis– monofactorial and multifactorial. The monofactorial analysis (see §3.2) focuses on only one factor at a time, while the multifactorial analysis (see §3.3) is based on all the levels and factors used in the data frame.

1. Synonyms of 'Disagree'
The most synonymous words of the verb disagree are - contradict, dispute, oppose and reject. The entries of these lexemes in the Oxford Dictionary are given from a-e.

a. disagree: have or express a different opinion; No one was willing to disagree with him.
b. contradict: deny the truth of (a statement) by asserting the opposite; The survey appears to contradict the industry's claims.
c. dispute: argue about (something); The point has been much disputed.
d. oppose: disagree with and attempt to prevent, especially by argument; A majority of the electorate opposed EC membership.
e. reject: dismiss as inadequate, unacceptable, or faulty; Union negotiators rejected a 1.5 per cent pay award.

Each of the above definitions relays close affinity with the verb under investigation, but the true semantic profile can be studied through the analysis done by BP in the following sections.

2. Monofactorial Analysis
Monofactorial analysis reveals the relation of one single level with the investigated lexemes. Monofactorial analysis can be executed in various ways. By using the VCD package in Rstudio, we can extract a barplot, association plot, mosaic plot and sieve plot in order to demonstrate synonyms-level relation.

The plots in Figure 1 and Figure 2 tell us an important story about the synonyms of disagree and the levels that I have used for tagging. From the barplot in Figure 1, it is understandable that all the synonyms have most entries in the Media genre. This is true for newspapers and magazines, which target to draw readers into building their opinions on something. It functions as a brain teaser. The corresponding association plot in Figure 2 asserts the figure drawn in the barplot. The over-representation of the lexeme dispute is visible in professional sites, while disagree has over representation in the Forum and oppose is over-representational in the Media genre. The synonyms disagree and oppose are under-represented in professional sites.
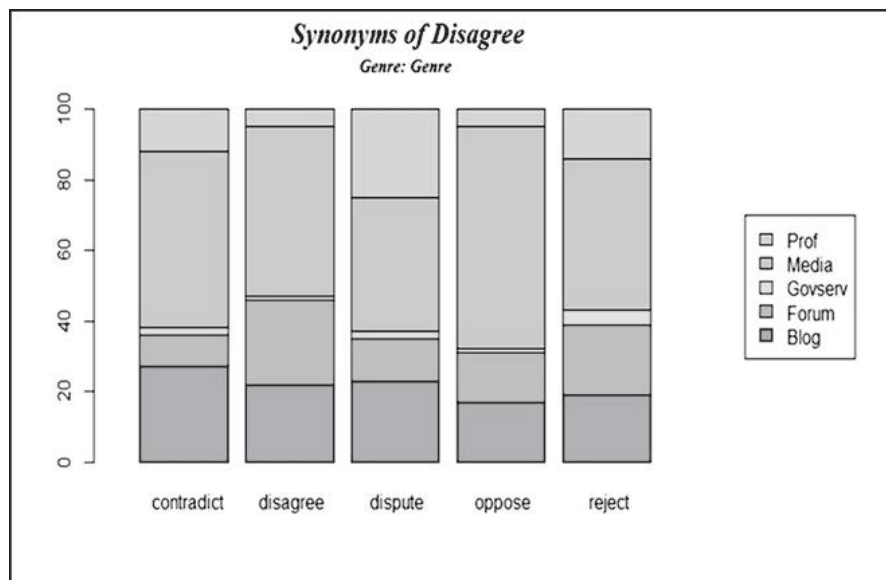
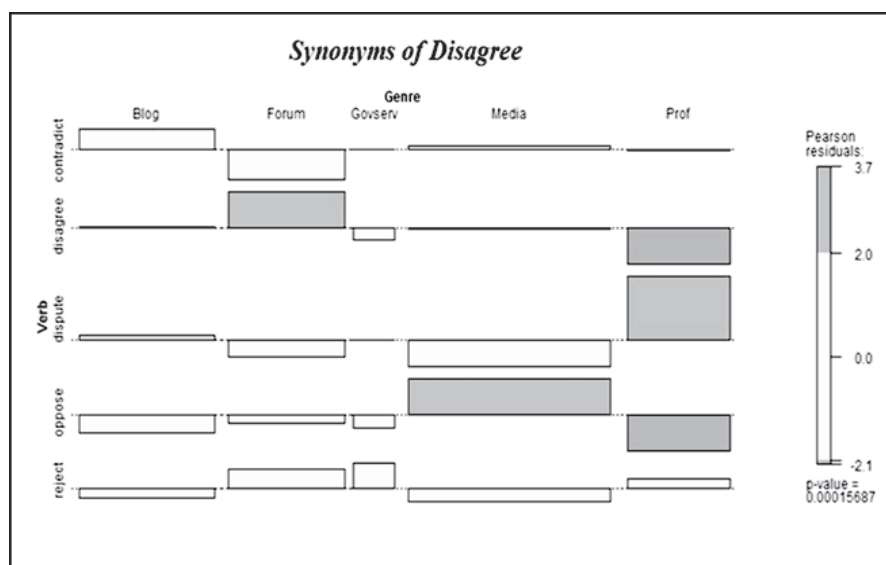**Figure 1: Barplot for the Genre factor and the synonyms**

Figure 2: Association plot for the Genre factor and the synonyms

The plots in Figure 3 and Figure 4 reflect the mono-level relation of verbs with the factor Verb_Form. Both the plots show us that the verb *contradict* has most instances in the present form, and is therefore over-represented in the given plots. The synonym *dispute* is more instantiated in the infinitive form, while *oppose* is over-represented in the past form. The possible reason for over-representation of *oppose* in past form could be due to the fact that all the other verbs can have stative nominal forms, while the noun of *oppose* is more related to an entity than to the nouns of states. Therefore, it is likely that *oppose* is more instantiated in the past form.
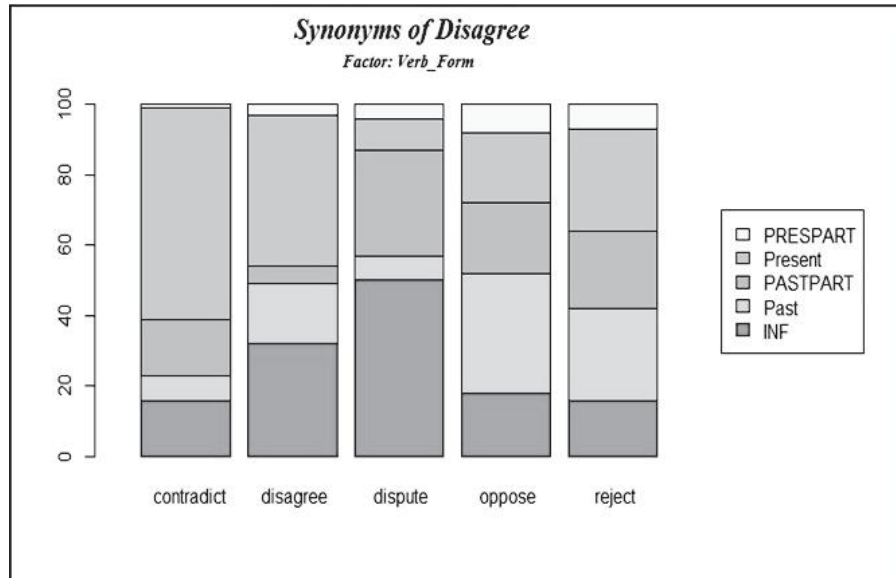
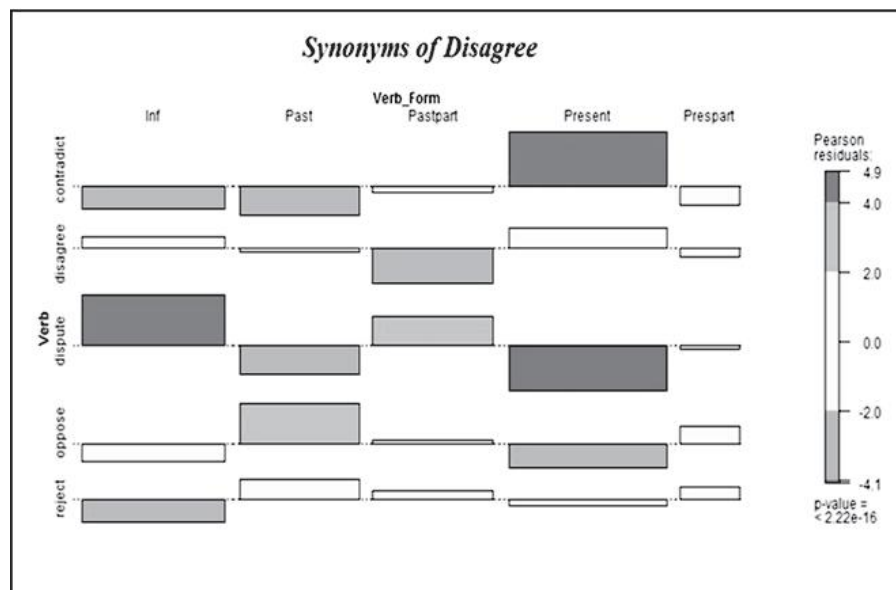**Figure 3: Barplot for the Verb_Form factor and the synonyms**

Figure 4: Association plot for the Verb_Form factor and the synonyms

Consider the sentences from Error! Reference source not found.-Error! Reference source not found.. If we look closely at sentence Error! Reference source not found., it seems that the lexeme opposition might represent two semantic meaning; one is the opposition party and the other is the opposition of something. Taking this inference into consideration, it is justifiable that the use of oppose is more instantiated in the past form. The semantic subject and semantic object are interesting factors in the sense that they show us the connection to

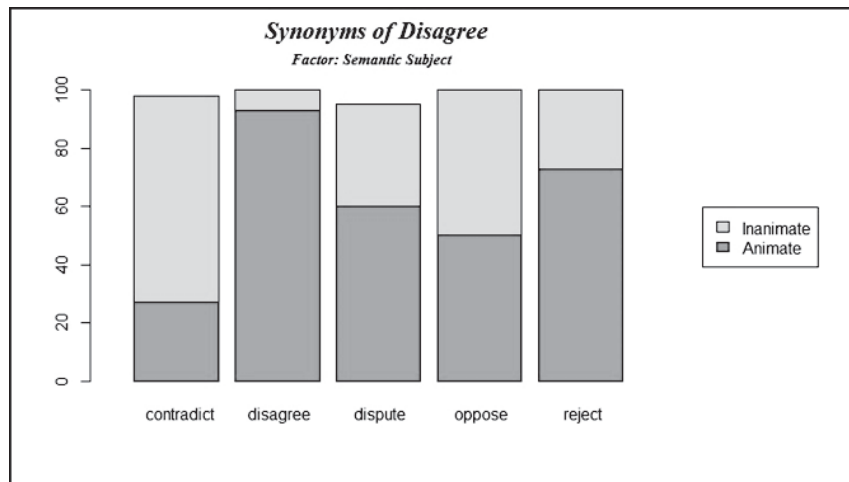animacy with the synonymous words. These two factors help us to show the words from a different perspective.

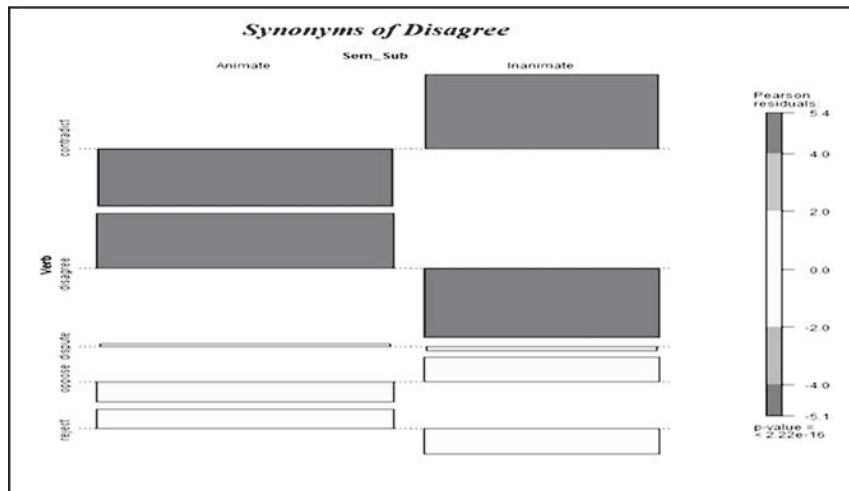**Figure 5: Association plot for the Sem_Sub factor and the synonyms**

Figure 6: Barplot for the Sem_Sub factor and the synonyms

When it comes to the semantic object that these synonymous verbs take, the plots in Figure 7 demonstrate that the only over-representation is for the verb disagree, which instantiates more animate objects than expected.
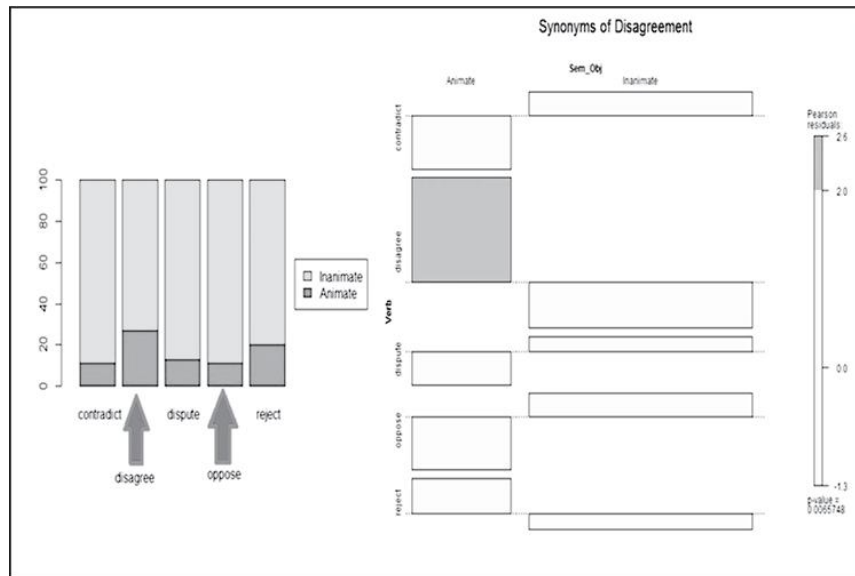
Figure 7: Barplot and Association plot for the Sem_Obj factor and the synonyms

3. Multifactorial Analysis

Monofactorial analysis gives us an overall picture of the lexemes based on the factors and levels that I have used to tag the different linguistic properties of lemmas. There are two types of results that we can extract using multifactorial analysis– two dimensional and multidimensional. In this paper, I am going to present multidimensional results through two different packages: configurational frequency analysis (cfa) and principal components analysis (pca).

Configurational Frequency Analysis (cfa)

CFA calculates the statistical importance of the co-occurrence of factor levels of our dataframe. This package shows us the result combining two or three factors. The more factors we will add, the less interesting the result will be. Thus, for each CFA analysis I have combined two factors so that I can extract an interesting analysis. I have combined the Genre and the Verb_Form factors, and the outcome is shown in **Error! Reference source not found.**

Table 4: CFA co-occurrence table for Genre and Verb_Form factor

			label	n	expected	Q	chisq	p.chisq	sig.chisq	z	p.z	sig.z
1	dispute	Prof	INF	16	3.2208	2.572410e-02	5.870416e+01	0.8864913	FALSE	7.1206852319	5.369039e-13	TRUE
2	reject	Forum	Past	11	2.8756	1.634279e-02	2.295377e+01	0.9999995	FALSE	4.7910098090	8.297201e-07	TRUE
3	oppose	Media	Past	23	8.8088	2.889140e-02	2.286238e+01	0.9999995	FALSE	4.7814624103	8.701227e-07	TRUE
4	contradict	Blog	Present	18	6.9552	2.240121e-02	1.753905e+01	1.0000000	FALSE	4.1879650108	1.407334e-05	TRUE
5	contradict	Media	Present	31	15.5848	3.182229e-02	1.524745e+01	1.0000000	FALSE	3.9847977739	4.715211e-05	TRUE
6	dispute	Blog	PASTPART	11	4.0176	1.407792e-02	1.213508e+01	1.0000000	FALSE	3.4835446135	2.474103e-04	TRUE

The marked rows reveal facts about the synonyms discussed in this paper. The results are also identical to the previously discussed monofactorial analysis. Row 3 in Table 4 shows that the verb oppose is used in the past form in Media for 23 instances. A few concordance lines from the corpus will help us to understand this statistic more conveniently.

   a. Jamaat-e-Islami is the country's largest Islamist party and it opposed Bangladesh's independence from Pakistan at that time. Some of…
   b. Last time, it was the political parties which opposed holding of upazila polls before the general elections. So the…
   c. …as he was not fully prepared. Prosecutor Rana Dasgupta opposed the defense counsel and argued that the submission was meant to…

The above concordance lines from a-c show that the verb oppose is used for the decisions which were already taken. We can also relate our previous assumption of using oppose as a verb in the past form. The verb contradict is instantiated in the present form in the Media genre, which was also reflected in the monofactorial analysis. The statistic in Table 5 shows the diplomatic function of this word along with inanimate subjects in Media.

Table 5: CFA co-occurrence table for Verb_Form and Sem_Sub factor

		label	n	expected	Q	chisq	p.chisq	sig.chisq	z	p.z	sig.z
1	contradict	Present	Inanimate	45	12.181081	6.825645e-02	8.842350e+01	4.877654e-12	TRUE	9.40337716	0.000000e+00	TRUE
2	dispute	PASTPART	Inanimate	24	6.683837	3.560680e-02	4.486188e+01	1.457240e-04	TRUE	6.69790088	1.057177e-11	TRUE
3	dispute	INF	Animate	41	15.396278	5.360872e-02	4.257851e+01	3.233276e-04	TRUE	6.52522103	3.395062e-11	TRUE
4	oppose	Past	Inanimate	19	7.113792	2.446295e-02	1.986029e+01	2.265793e-01	FALSE	4.45648812	4.165661e-06	TRUE
5	disagree	INF	Animate	32	16.206609	3.312418e-02	1.539871e+01	4.962298e-01	FALSE	3.92309955	4.370850e-05	TRUE
6	disagree	Present	Animate	37	19.821929	3.630361e-02	1.488685e+01	5.329408e-01	FALSE	3.85834838	5.787793e-05	TRUE
7	reject	Past	Animate	23	11.344626	2.419858e-02	1.197463e+01	7.457241e-01	FALSE	3.46043841	2.696483e-04	TRUE
8	dispute	Present	Inanimate	1	11.808113	2.246113e-02	9.892800e+00	8.721662e-01	FALSE	3.14528223	8.296334e-04	TRUE
9	reject	PRESPART	Inanimate	6	1.797991	8.554543e-03	9.820335e+00	8.758408e-01	FALSE	3.13374131	8.629647e-04	TRUE

The genre of Media uses inanimate subjects in the present form to avoid any direct confrontation with readers. That is, the media talks about things rather than humans. Moreover, oppose can use animate subjects (sentences from a-c), as the action has already taken place in the past. Table 5 also shows that dispute and disagree are instantiated with animate subjects in the present and infinitive forms. We can see a clear distinction between disagree and contradict; disagree is semantically used with an animate subject and contradict with an inanimate subject.
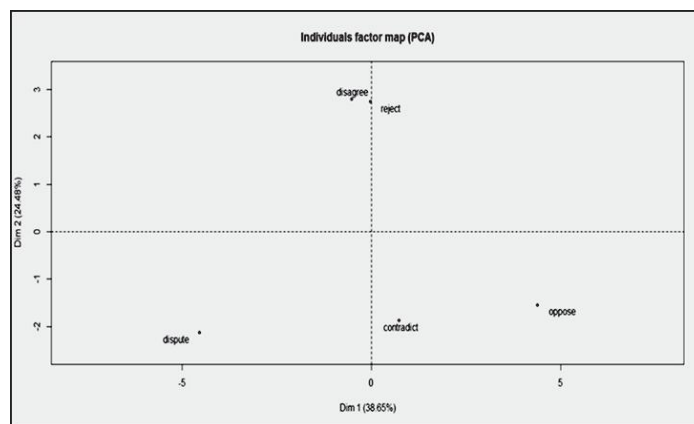
Table 6: CFA co-occurrence table for Sem_Sub and Sem_Obj factor

		label	n	expected	Q	chisq	p.chisq	sig.chisq	z	p.z	sig.z
1	contradict	Inanimate	Inanimate	70	31.639957	0.0831455676	46.587422648	1.931166e-09	TRUE	6.81963508	4.563572e-12	TRUE
2	disagree	Animate	Animate	26	9.973298	0.0331797442	25.754289315	3.546723e-05	TRUE	5.07486840	1.938821e-07	TRUE
3	contradict	Animate	Inanimate	17	50.457406	0.0756026790	22.185008940	1.841208e-04	TRUE	4.71009649	1.237998e-06	TRUE
4	disagree	Inanimate	Inanimate	6	32.285671	0.0570541638	21.400716541	2.636894e-04	TRUE	4.62609085	1.863159e-06	TRUE

Table 6 also confirms the abductive inference that Media uses the inanimate subject and object significantly in the present form in order to avoid any kind of debate and to present neutrality of opinions.
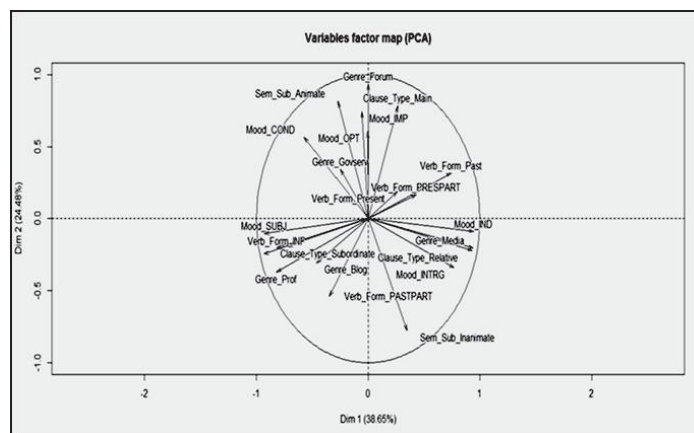
4. Principal Component Analysis

For the Principal Component Analysis (henceforth, PCA), I have used the FactoMineR package in Rstudio. The PCA performs with supplementary individuals, supplementary quantitative variables and supplementary categorical variables. The PCA map in Figure 8 shows the position of the synonyms in a two dimensional graphs based on the tags I have used.

Figure 8: Individual factor map

The PCA map in Figure 8 illustrates that on the X-axis, contradict, oppose and reject are on the same side of the graph while disagree and dispute are on the opposite side. On the Y-axis, dispute and reject are on the opposite side of the other synonyms. This map is based on the closely shared levels by the synonymous verbs of disagree. The previous sections show that at least for some levels contradict and disagree are different semantically, which is also represented in the PCA map. The variables factor map is more detailed in reflecting the levels which put certain lexemes on the same axis of the graph. The following variables factor map in Figure 9 shows us the significant variables by long arrows which are representations of the commonly shared levels among the investigated lexemes.

Figure 9: Variables factor map

The variables factor map in Figure 9 should be read in relation to the two dimensional map presented in Figure 8. The long lines of levels for the verbs contradict and oppose are Sem_Sub_Animate, Mood_INTRG, Genre_Mood and Mood_IND. Thus, contradict and oppose share these common levels. The significant factors and levels for the verb disagree are Sem_Sub_Animate, Mood_COND and Genre_Forum. These readings are identical to the previous analysis in this paper.

Findings

Based on the analysis given above, it is clear that using the BP approach, we find out more than our intuitions. We have seen that the synonyms of disagree slightly differ in the real usage. Speakers' selections of the synonyms are based on different sociolinguistic factors, and these factors are not easily foreseen. The lemmas of contradict and disagree reflects interesting facts, which are consistent throughout different analysis presented in this paper. Not only do these two verbs but also the verb *oppose* shows us insightful semantic properties. I have summarized the findings of this investigation from a-e.

a. Professional sites display more instantiations of the verb *dispute* in the infinitive form than in any other genres and forms.

b. The verb *oppose* is found to display more instances in past form in Media. The possible motivation for this reveals that other synonyms can take the nominal forms to be used in the past without any ambiguity in semantic representation.

c. The verb *disagree* is found to have significant percentage of animate subjects and the highest percentage of usage is in the genre of Forums. It is plausible that Forums have a lot of direct human interactions. Thus, the websites which belong to Forums instantiate the verb *disagree* more often than other synonyms.

d. The verb *contradict* is used in sentences where the subject is mainly inanimate. This verb is used in Media in present forms, which shows the effectiveness of the word being used as neutral in strictly pragmatic sense. It also does not reflect any direct confrontation and debate with any living entity. Hence, the use of this verb with inanimate subjects in Media is well justified. *Contradict* has also a significant percentage of combination of inanimate subjects and inanimate objects.

e. The verbs *dispute* and *reject* are also found significantly with animate subjects in the websites which belong to Forums and Blogs, where there is a lot of human interactions.

Conclusion

The Behavioral Profile approach helps us understand the semantic differences of the near-synonymous or synonymous words. The recent work in this area includes discussion of polysemys and synonyms. I have focused on the synonyms of *disagree* in this paper to investigate the linguistic differences that they bear in the real context within the GloWbE corpus. This corpus-based distributional approach to study near-synonymous words is a great tool, which can reveal the sociolinguistic conventions attached with the lexemes. This study is based on 100 concordance lines for each synonymous word from the corpus. A larger dataframe will restate the facts revealed in this paper.

References

Dobric, N. (2010). Word Sense Disambiguation Using ID Tags – Identifying Meaning in Polysemous Words in English. *In Vitas D. Krstev C. (Eds), Proceedings of the 29thInternational Conference on Lexis and Grammar/LGC, 97-105.* University of Belgrade.

Firth, J.R. (1957). *Papers in linguistics 1934-1951.* London: Oxford University Press.

Gries, S. Th. & Otani, N. (2010). Behavioral profiles: a corpus-based perspective on synonymy and antonymy. *ICAME Journal 34*, 121-150.

Gries, S. Th. (2012). Behavioral Profiles: A fine-grained and quantitive approach in corpus based lexical semantics. *In Garry Libben (Eds), Methodological and Analytic Frontiers in Lexical Research.* Amsterdam: John Benjamins Publishing Company.

Gries, S. Th., & Divjak, D . S.. (2009). Behavioral profiles: a corpus-based approach towards cognitive semantic analysis. *In V. Evans & S. S. Pourcel (Eds.), New directions in cognitive linguistics*, 57-75. Amsterdam: John Benjamins Publishing Company.

Han, W., Antti A., & Newman, J. (2013). Topic marking in a Shanghainese corpus: From observation to prediction. *In W. Stefanie (Ed), Corpus Linguistics and Linguistic Theory,* 1-29. Berlin: Mouton De Gruyter

Kakava, C. (2002). Opposition in Modern Greek discourse: cultural and contextual constraints. Journal of Pragmatics 34.10-11, 1537-1568.

Koczogh, H. (2013). Scrutinizing the Concept of (Verbal) Disagreement. *Argumentum 9*, 211 222.

Kotthoff, H. (1993). Disagreement and Concession in Disputes: On the Context Sensitivity of Preference Structures. *Language in Society 22.2*, 193-216.

Newman, J. (2011). Corpora and Cognitive Linguistics. In S. Th. Gires (Ed). Special issue on Corpus studies: future directions. *Brazilian Journal of Applied Linguistics 11.2*, 521-559.

Xiao, R., & Mcenery, T. (2006). Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective. *Applied Linguistics 27.1*, 103-129. Oxford University Press.