Research on Humanities and Social Sciences
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol.4, No.3, 2014

An Experiment in Plagiarism Detection in Academic Research Articles Using Attributional Techniques

Khalid Shakir Hussein

English Department, Thi-Qar University, Iraq

E-mail of the corresponding author: khalidshakir74@gmail.com

Abstract

There are certain overlapping aspects that brings plagiarism detection and authorship attribution together in one basket. Suspected cases of plagiarism can be interpreted as special cases of disputed or misattributed authorship. Being so, the same techniques used to resolve doubtful cases of attribution can be used to investigate any potential existence of plagiarism. Principal components analysis and cluster analysis (henceforth PCA and CA) are among the popular statistical techniques used to proceed with various attributional scenarios. These two techniques are used throughout this paper to explore the patterns of function words displayed in seventeen samples of one specific *genre* (academic research articles).

A survey is conducted over various cases of academic attribution: academic writing in English as a First Language and as a Second Language, and even cases of research articles with mixed authorship. Function words have been targeted in this paper as possible indicators of the author's identity. Accordingly a set of English function words is tested using WordSmith Tools (version 5.0). It turned out that the multivariate techniques (represented by PCA and CA) are most likely robust for addressing the type of issues raised about plagiarism and authorship attribution. Besides, it appeared that the statistical patterns of function words usage are rather relevant markers to deal with various scenarios of potential cases of plagiarism. This could explain the three different clusters plotted in the data environment for Halliday's samples, the Phillippine's samples together with his collaboratively authored ones, and the Iraqi's suspected samples that represented a highly potential case of plagiarism.

Keywords: Authorship Attribution, Plagiarism Detection, Authorship Verification

1. Introduction

The concept of plagiarism is quite controversial and has recently raised interest in most universities all over the world. It is stunning how many definitions have been suggested to set up the underlying scaffold of the plagiarizing behavior. Nonetheless, one cannot come across a single comprehensive definition without causing any harms or misinterpretations to the conceptual essence of plagiarism. This is not surprising in the light of the diversified nature of plagiarism as a practice. Turell (2004:4), for example, suggests a rather macro-definition with far-reached identifications: plagiarism comprises any attempt involving "intentional lifting of an idea and/or intentional copying of the text (linguistic, musical, etc. . .) used to express that idea, to cover up non-originality." Certain questions about the scope of plagiarism are clearly triggered by this definition. It underscores the techniques used, (*lifting of an idea- copying of the text*), and describes both as "intentional", but it does not say much about the range of what is lifted or copied.

However, the definition used in the American Association of University Professors sounds more articulate and explicit (cited in Roig, 2006: 3): " . . . taking over the ideas, methods, or written words of another, without acknowledgement and with the intention that they be taken as the work of the deceiver."

Coulthard and Johnson (2007) managed an *educationally* biased definition and it is officially published on the University of Birmingham website:

"PLAGIARISM AND CHEATING IN EXAMINATIONS

Plagiarism is a form of cheating in which the student tries to pass off someone else's work as his or her own. … Typically, substantial passages are 'lifted' verbatim from a particular source without proper attribution having been made. To avoid suspicion of plagiarism, students should make appropriate use of references and footnotes."

(University of Birmingham http://artsweb.bham.ac.uk/arthistory/
declaration_of_aship.htm [accessed 22 October 2013])

There is still a long list of running definitions that hang over the heads of investigators and scholars interested in this area of concern. A sort of consensus, however, among researchers came to highlight *the intentional* perception of plagiarism cutting across most of these pertinent definitions. A matter which might go with a context of legal accusations and judicial indictments that presume intentionality from the defendant's side.

Simply speaking, plagiarism is seen as an *offence*, especially in those countries whose legal systems acknowledge *Intellectual Property Laws* (United States, Australia, Great Britain, Canada, . . .etc.). In such countries, plagiarism is seriously tackled and linguists with appropriate expertise are regularly asked to investigate and have their say on controversial cases of plagiarism. However, in the *Iraqi Civil Law* item number (40) for 1951(see **www.iraqilawconsultant.com**), plagiarism is mentioned as a sort of ethical misconduct and irresponsible use of the intellectual property. But this law is still inactive and not taken seriously and no case of plagiarism was registered in the history of the Iraqi courts, not to mention the history of the Iraqi universities.

The researcher thinks that plagiarism is an inherent area of concern that should be expounded under a particular type of authorship misconduct or *misattribution* especially within the academic context. Plagiarism can even be recognized indirectly along any attempt to classify authorship types. Love's classificatory scheme (2002) stands out as one of the most explicit indications of plagiarism as a special verifying case of authorship attribution.
Moreover, plagiarism is not only a particular practice of authorship verification but even a case in which authorship is *intentionally* misattributed. Coulthard (2005:43) recognizes plagiarism as "one major authorship detection problem". In spite of this perspective what is misattributed or unacknowledged is the name of the original author and not the distinctive *authorial markers* which are still lurking behind the plagiarized textual body of the authentic author. Such markers are supposed to be left intact waiting for an attentive analyst to dig them up. These markers need to be genuine so that they can bear the fingerprints of the real author whose text has been plagiarized. The notion of genuine markers entails a plausible question about the kind of markers that distinguish the author's original contribution and discriminate him from all the other alleged authors.

One of the most evident overlapping areas between authorship attribution and plagiarism is to establish the linguistic markers that might be crucial and determinative in detecting plagiarism. Two basic principles, used in sociolinguistics and pragmatics, are identified as being pertinent to the linguistic description of plagiarism:

" . . . whenever a speaker or a writer produces a message, he or she will produce a *unique* and *idiosyncratic* text with a number of linguistic 'markers' or 'resources' that will make it *unrepeatable*."
" . . . speakers and writers . . . pay little attention to the form that either speech or writing take; they are therefore not aware of those specific linguistic 'markers' and 'resources' . . ., and consequently *if these 'markers' remain unnoticed by the authors themselves they will remain unnoticed by any user who would try to plagiarize, imitate or copy them.*"

(Italics mine,Turell, 2004: 7-8)

These two principles have different wordings but similar content to those used in establishing attribution: first, linguistic uniqueness or individuality of speakers and writers, and second, the unconscious use of given discriminative authorial markers, a matter which lies beyond the plagiarist's conscious control. But what kind of markers do such underlying principles govern?

There is a whole series of criteria and linguistic resources and discourse strategies suggested by linguists to capture the authorial blueprint: *the degree of unity*, *completeness and coherence/cohesion* of the texts under comparison, *inconsistency in referential style*, *decontextualization*, . . . etc. (see Turell, 2008: 282-7). However, the researcher is particularly interested in one particular marker: it is the usage of function words which has a long standing history since Mosteller and Wallace's analysis of the twelve disputed Federalist Papers (1964). Ever since their statistical analysis of (30) *function words* extracted from the various texts of Federalist Papers, the latter became a touchstone for the credibility of the new methods of authorship investigation.

1.1 Function Words

The use of function words is still considered the essence of a popular and successful method performed to characterize a specific author in terms of the way he uses such words throughout his writings. The appeal of

function words in attributional studies lies in their being significant and stable indicators of authorial identity. But one may wonder about the rationale behind the assumption that people tend to express themselves in stable and unique patterns of function words usage.

Much has been written about this rationale which almost always instigates *three* salient characteristics about function words: *first*, due to their high frequency in the language, function words are used in preference to others (Argamon and Shlomo, 2005). *Second*, the low "semantic load" of these words together with their highly grammaticalized roles make their frequencies rather stable regardless of the topic of the text (Koppel, et al., 2007:9). This very particular characteristic empowers the researchers to attribute the texts of different topics to the same author. The *third* characteristic was brought up by psycholinguistics: it was found that function words usage lies beyond our conscious control. Therefore, authors can not keep this use under their direct control. Friederici (1996, cited in Hoover, 2001: 422) presents a strong neurophysiological evidence on the speed and location of the processing of closed-class (function words) versus open-class (context words) within the brain. It was found that after the age of ten speakers tend to process closed-class words more rapidly and in a different area of the brain than open-class words (ibid.). This rapidity of processing contributes greatly in the automation of closed-class words usage. Thus, it is the highly automatic processing of function words that enhances the possibility of an author "word print" that could be used actively in resolving attributional doubts.

It is therefore interesting to utilize function words as highly reliable markers of style giving unavoidable allusions to authorial attribution. In English, a well-known collection of about 70 function words is often used for this purpose. However, in attributional studies function words are defined in a way that sounds rather broad including *prepositions*, *conjunctions*, *articles*, *auxiliary* and *modal verbs*, *determiners*, or even elements such as words describing *quantities*, *degree adverbs, numbers* and *interjections*.

They are all included in the lists of function words used in the typical modern attributional studies. These words might essentially be topic-independent, though they are not, strictly speaking, classified as function words in the very sense of this grammatical term. Consequently, the precise choice of function words is not always crucial. What matters, however, is the *low semantic load* of the linguistic items in question regardless of their usual grammatical classification.

2. Analysis Procedures

The corpus used throughout this paper was compiled via the Internet. As for the digital samples included in the corpus, they will be subjected to *six* analysis procedures:

1. *Authenticity Investigation*: it is quite expected that "a corrupt sample" would most definitely produce "a corrupt analysis". Consequently, it should be determined that the digital samples selected for authorship analysis are clean samples, a task which sounds extremely difficult if not impossible (Juola, 2008:247). Since the samples selected for this study are machine-readable, the scanning or retyping processes could be a very threatening source of all types of errors. The researcher tried his best to check the authenticity of each sample making sure that each one is highly representative of the authors involved. Whenever there are hard copies of the texts they should be compared to their digital ones. This process might appear boring and painstaking but it is inescapable. Moreover, there are certain "non-authorial" materials that should be removed from the samples: *major heads*, *section heads*, *page numbers*, *quotations*, and so forth. They could be a severe threat to the statistics ascribed to the author's linguistic habits.
Hence, only the main body of the texts will be considered: *titles, author names, dates, . . .*etc. all were excluded. After all, every sample should be authenticated in a way that sounds independent and reliable. Otherwise, the sample would be eliminated for its potential extraneous variables that might influence the statistical results
2. Transcribing digital samples into *plain text format*
3. Grouping all the samples into one master corpus
4. Analyzing samples with their master corpus via *WordSmith Tools (5.0)* for function words frequency and word count, besides producing some sort of charts representing basic statistical descriptions.
5. Importing WordSmith Tools *outputs* into an excel spreadsheet in a form of matrix.
6. Conducting a thorough statistical analysis to the matrix using *SPSS (14.0)* in general, and *PCA* and *CA* in particular.

Below is a description of the samples surveyed in this paper:

Table 1. Corpus Description

Author	Text-samples	Samples Number	Genre
M.A.K.K Halliday	(1,000-word) samples selected from different academic articles written between (1967) and (2002).	11	Research Article (Linguistics)
A Philippine Researcher	(1,000-word) samples (2010)	2	Research Article (Linguistics)
An Iraqi Researcher	(1,000-word) samples (2005)	2	Research Article (Linguistics)
Mixed Authorship (the Philippine and a Native Researcher)	(1,000-word) samples (2012)	2	Research Article (Linguistics)

3. The Experiment

Since the researcher looks at plagiarism as a particular case of authorship misattribution, the experiment conducted along this section is performed to verify the validity of such a perspective. Two multivariate methods (PCA and CA) will be used to go through the plagiarism corpus described in Table (1). Plagiarism corpus consists of 17 academic research articles: *eleven* are attributed to Halliday and were written over a long span of time (four decades), *two* are attributed to a Philippine researcher, and other *two* to an Iraqi researcher. As for the *two* mixed authorship samples, they were written collaboratively by the same Philippine researcher involved in this corpus and another native English researcher. It should be noted that the Philippine and Iraqi researchers are non-native English second language authors. The researcher avoided using their names to escape the problems of taking the necessary permissions or what is usually called *Copyright Clearance*.

The distribution of the samples was set for a purpose. First, Halliday's eleven samples represent an opportunity to figure out whether there are any irregular shifts or discontinuities in his writing style after four decades of writing academic research articles. Second, it is quite crucial for researchers working on plagiarism in the academic setting to find out about the academic authors' stylistic capabilities. Those writers engaged in academic second language writing are particularly crucial in this concern. Do they show any capability of establishing their own independent authorial identities throughout certain linguistic habits related to the way they use function words? Or are they intrigued in a sort of patchwriting? Third, it is quite common to have research articles of mixed authorship (two or more authors are involved). Does that have any repercussions on identifying changes of authorship? Is it possible to specify which part of a research article was written by which researcher?

The corpus-based analysis of such a database is supposed to illuminate some implications about the nature of *academic attribution* and the significance of the concept of statistical stylistic consistency in alerting any signs of potential academic plagiarism. Moreover, the validity of using function-word patterns in authorship attribution will be stretched to its furthest limits by being retested on a rather different body of data controlled by one specific *genre* and *topic*. This type of data is related to the language used in the academic contexts focusing on some particular academic topic (linguistics). Do function words show any systematic co-occurrences in the academic language? Is it possible to interpret academic attribution in terms of the *existence* or *absence* of such hypothetical patterns of co-occurrences?

A wordlist was obtained by *WordSmith Tools* (Version 5.0) figuring out the top 28 function words in the master corpus. Due to the scarcity of the textual data in the research articles under investigation, especially after removing every citation, major and minor titles, function words holding zero-frequency only in one specific sample were included in the frequency list. Other words showing zero frequency in two or more samples were excluded from the list. Table (2) tabulates the commonest 28 function words surveyed throughout the master corpus.

Table 2. The Top 28 Function Words in Plagiarism Corpus

WordSmith Tools -- 30/6/2013			
N	**Word**	**Freq.**	**%**
1	THE	1092	6.579014301
2	OF	769	4.525301933
3	AND	567	3.042348146
4	IN	504	2.850619221
5	TO	467	2.741680622
6	A	341	1.941599131
7	IS	295	1.885542512
8	THAT	255	1.472761512
9	IT	222	1.191325483
10	AS	221	1.189229405
11	THIS	175	0.891810656
12	ARE	168	0.856138229
13	FOR	129	0.657391846
14	NOT	122	0.62171942
15	BE	117	0.59623909
16	WITH	108	0.550374568
17	BUT	100	0.509606063
18	WHICH	98	0.499413967
19	ON	90	0.458645463
20	HAS	90	0.458645463
21	AT	88	0.448453337
22	WHAT	82	0.417876989
23	FROM	81	0.412780911
24	THERE	77	0.392396688
25	SO	74	0.377108485
26	OTHER	61	0.31085971
27	CAN	57	0.290475458
28	ALL	44	0.244015381

The 28 function words above account for 36.31% of the total words in the major corpus. More than 25 function words were excluded from the list for their nil frequencies that recur in more than one sample. This could explain the slightly low percentage of the 28 words.

3.1 Statistical Analysis Through PCA

Carrying out the PCA, the researcher framed the variables behavior between the first two principal components. The first component explains 32.17% of the total variance. The variation on the second principal component constitutes 13.85% of the variance. This makes the total variance captured by the two: 46.02%. Table (3) below highlights the total variance explained by the first two principal components.

Table 3.Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.007	**32.170**	32.170	9.007	32.170	32.170
2	3.880	**13.859**	46.028	3.880	13.859	46.028
3	3.662	13.078	59.107			
4	2.385	8.517	67.623			
5	1.958	6.993	74.617			
6	1.633	5.832	80.448			
7	1.422	5.078	85.526			
8	.921	3.290	88.817			
9	.771	2.753	91.569			
10	.692	2.473	94.042			
11	.489	1.748	95.790			
12	.414	1.480	97.270			
13	.362	1.293	98.564			
14	.215	.769	99.332			
15	.120	.429	99.761			
16	.067	.239	100.000			
17	1.47E-015	5.26E-015	100.000			
18	4.90E-016	1.75E-015	100.000			
19	4.38E-016	1.56E-015	100.000			
20	3.40E-016	1.21E-015	100.000			
21	1.13E-016	4.03E-016	100.000			
22	4.16E-017	1.48E-016	100.000			
23	-2.55E-017	-9.10E-017	100.000			
24	-1.24E-016	-4.42E-016	100.000			
25	-2.04E-016	-7.28E-016	100.000			
26	-3.42E-016	-1.22E-015	100.000			
27	-5.77E-016	-2.06E-015	100.000			
28	-6.41E-016	-2.29E-015	100.000			

Figure1. plots the 28 function words in the scatterplot below displaying rather interesting patterns of three isolated combinations of function words, leaving *in* down by itself. On the far left of the first principal component, the function words *of, for, and* dominate the variation range, whereas *it, at*, *that* correlate closely with the variation on the far right of the same component. The top of the second component is dominated by *on*, *from*, *of* and the function words *which, in, is* reside at the bottom. Table (4) explains the numeric values of the variance conditioned by the 28 function words on both components in a form of matrix.

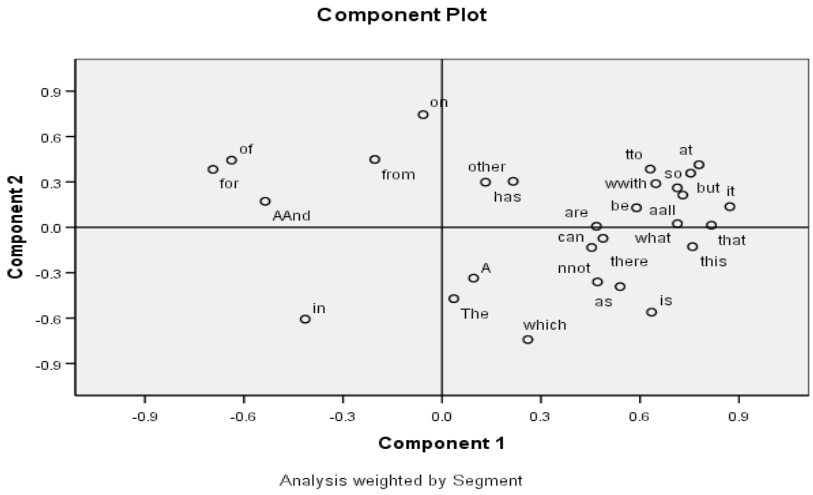

Figure 1. Component Plot of Function Words

Table 4. Component Matrix of the Variance

	Component	
	1	2
The	.036	-.472
of	-.**638**	.**444**
and	-.**536**	.172
in	-.414	**-.607**
to	.631	.385
a	.096	-.336
is	.635	**-.560**
that	.**816**	.014
it	.**872**	.137
as	.540	-.392
this	.759	-.128
are	.468	.007
not	.471	-.361
be	.590	.129
for	-.**694**	.383
but	.713	.261
with	.648	.289
at	.**779**	.414
which	.260	**-.742**
has	.132	.298
on	-.057	.**744**
what	.713	.024
there	.488	-.073
so	.754	.358
from	-.204	.**449**
other	.216	.304
can	.454	-.133
all	.730	.213

Extraction Method: Principal Component Analysis. 2 components extracted.

The behavior of the function words plotted above still of no use to plagiarism questions since what matters in this context is the influence these words have over the behavior of the seventeen samples.

The scatterplot in figure (2) reveals more interesting and explicitly readable plots of the individual samples. The scatterplot offers quite insightful responses to plagiarism issues. Though Halliday's eleven samples reveal somewhat diverse plotting, they tend to cluster in the upper right occupying a wide area of the data environment. The Philippine's samples come so close to each other residing in the upper left. As for the Iraqi's plots, sample fifteen looks more discriminated from Halliday's. It fairly detaches itself to settle in the lower right of the figure. But sample fourteen, in contrast, is most strongly identified with Halliday's cluster. This raises a red flag that alerts a likely case of authorship misattribution and provokes doubts of plagiarism. For samples sixteen and seventeen, it can be noted that the sixteenth sample resides at the Philippine's authorial territory (upper left) suggesting that the Philippine's major authorial contribution is located in that part of the research article. Whereas the seventeenth sample occupies an independent plot that comes closer to the center of the figure. It is unattributable and does not belong to any of the clusters identified so far. It even does not hold any affinities with the Philippine's corner. Thus, it seems that this sample was penned by a different writer whose patterns of function words are discriminated from all the three authors involved in this corpus.

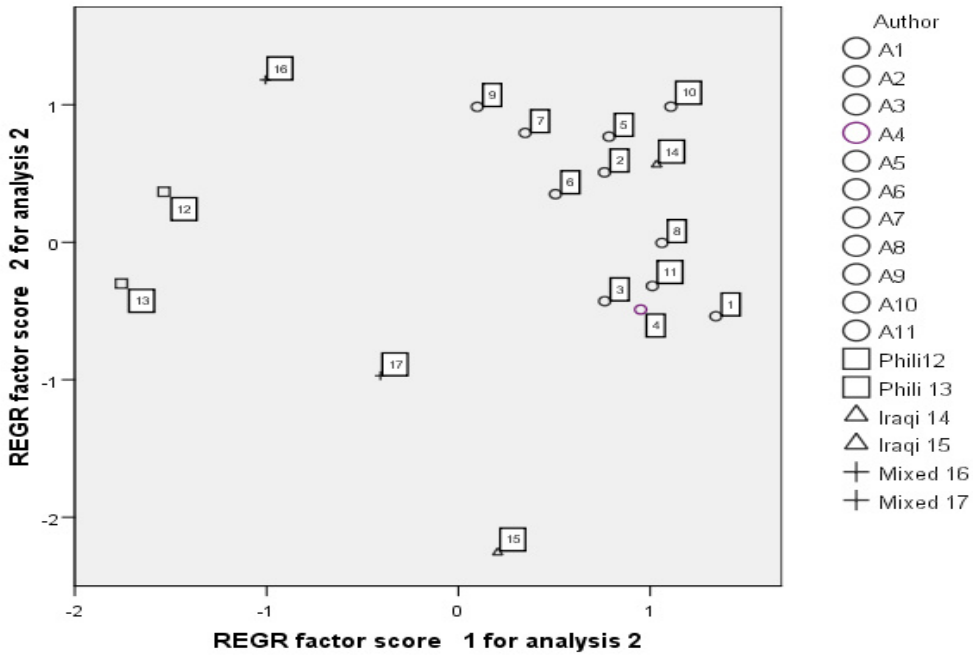

Figure 2. Seventeen Samples Plot

3.2 Statistical Analysis Through CA

For the doubts of plagiarism to be sufficiently cleared, more confirmations for the outputs produced by PCA are most definitely welcomed. Here comes the role of CA to sustain or undermine the preliminary insights the researcher has achieved so far about the seventeen samples plotting.

The dendrogram shown in Fig 3. translates the findings of PCA in a rather interesting way. Three distinct clusters are identified. The first cluster coincides with the highest split (the rescaled distance is 25), combining the Philippine's two samples (twelve and thirteen) on the third vertical line together with sample sixteen lagging behind. Thus, the textual similarity assigns the sixteenth sample to the Philippine's cluster, though it is still an outlier. The second cluster belongs to the second split with two branches: one holding for the Iraqi's fifteenth sample and another for the mixed authorship sample. Though they belong to one cluster, their similarity is quite questionable as they are identified on the fifth vertical line. Then, Halliday's eleven samples start clustering on the third split that brings them together in one cluster. All the samples, except ten, agglomerate into one group with an intensive degree of similarity as they all cluster on the first vertical line. The plagiarism doubts are

evidently arisen and underscored in the fourteenth sample. This sample is recognized most definitely with samples (2, 5, and 6) on the first vertical line indicating a high degree of textual similarity.

This would really cast serious suspicions about the authentic authorship of the fourteenth sample. Such suspicions might be pursued by following the traditional way of digging up the references used in the research article. It turned out that the sample under consideration was an introductory part of the article and related to Halliday's systemic functional grammar. Whole passages were used freely (word for word) by the Iraqi researcher without making any proper attribution to indicate the sources they have been cited from.

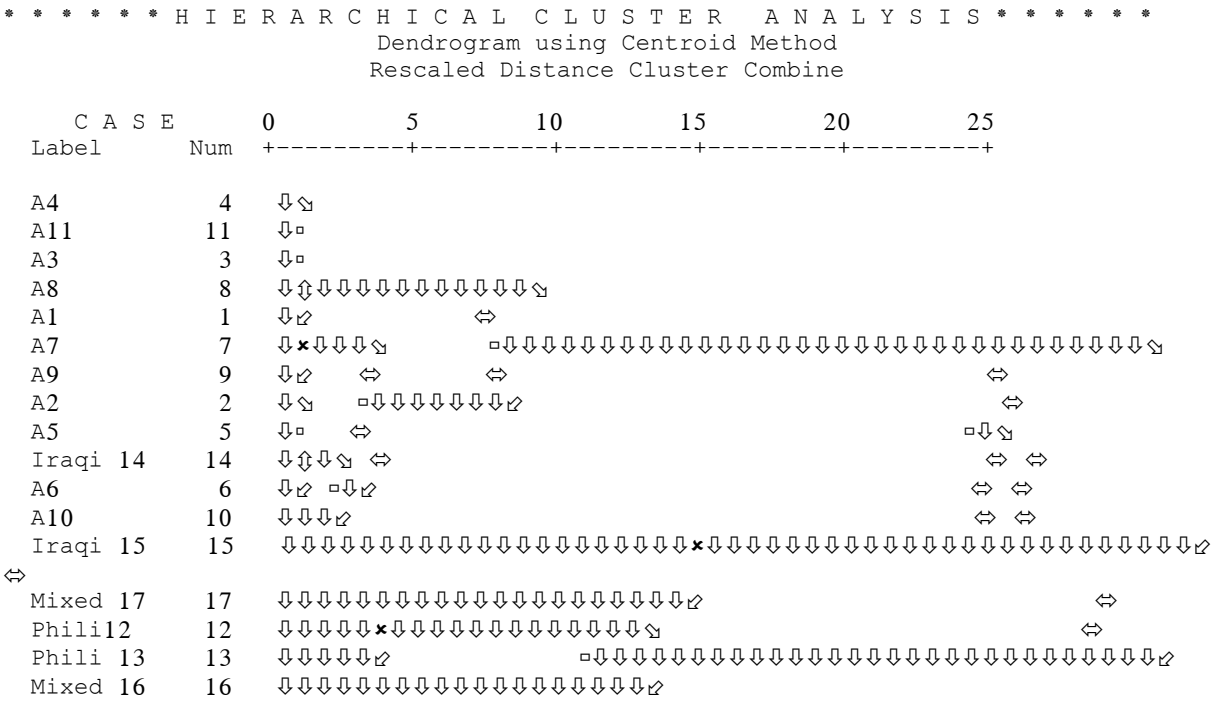

Figure 3. Hierarchical Cluster Analysis

4. Conclusions

Our experiment has shown a clear overlapping area between plagiarism detection and authorship attribution. The techniques used to attribute authorship can be useful for alerting researchers to potential cases of suspected plagiarism. PCA and CA turned out to be robust in resolving problems of academic attribution even with cases of mixed authorship. It should be remembered, insofar as PCA and CA are concerned, that this experiment is based on a simple yet plausible rationale: any potential case of plagiarism would float up on the statistical environment in a form of overlapping samples with no clear area or cluster to be reserved for any specific academic researcher.

The research articles of the Philippine (samples 12, 13, 16) have shown authentic statistical fingerprints of their own with reasonable textual similarity. The plagiarism doubts about the Iraqi's research article were triggered by the extreme textual similarity that brought the fourteenth sample and Halliday's samples together in one cluster. Consequently, these two statistical techniques proved reliable and relevant to resolve cases of authorship attribution and plagiarism checkups in the academic contexts.
Moreover, the experiment revealed consistent statistical behavior observed along the overall way the seventeen samples cluster. It is extremely evident that they are indisputably distributed into three approximately separated authorial boundaries with two skew samples: one suggests a suspected plagiarism, and the other is an unattributable sample. Therefore, a judgment can be worked out in favor of *a positive attribution* of only fifteen

samples to the three academic researchers and *negative* or *disputed attribution* of two samples: one claimed by the Iraqi researcher and the other by an anonymous researcher.

References

Argamon, S., and Levitan, S. (2005), Measuring the usefulness of function words for authorship attribution. [Online] Available: http://SArgamon,SLevitan-ACH/ALLC,2005-tomcat-stable.hcmc.uvic.ca (May 19, 2005)
=
Coulthard, M., & Johnson, A. (2007) *An Introduction to Forensic Linguistics. Language in Evidence*. London/New York: Routledge

_____,(2005). The linguist as expert witness. *Linguistics and the Human Sciences*, 1(1), 39-58

Hoover, D. I. (2001). Statistical stylistics and authorship attribution: an empirical investigation. *Literary and Linguistic Computing,* 16(4), 421-444

Koppel, M., Schler, J. and Shlomo, A. (2007), Computational methods in authorship attribution. [Online] Available: http://u.cs.biu.ac.il/~koppel/papers/authorship-JASIST-final.pdf (July 12, 2007)

Love, H. (2002). *Attributing Authorship: An Introduction*. Cambridge: Cambridge University Press.

Mosteller, F. and Wallace, D. (1984). Inference and Disputed Authorship: The Federalist. (Reading, MA: Addison-Wesley Pub. Co., 1964); 2nd edition as *Applied Bayesian and Classical Inference: The Case of the 'Federalist' Papers*. New York: Springer-Verlag.

Roig, M. (2006). Avoiding Plagiarism, Self-Plagiarism, and other Questionable Writing Practices: A guide to Ethical Writing. [Online] Available: http://facpub.stjohns.edu/~roigm/plagiarism/index.html

Turell, M. (2004), Textual kidnapping revisited: the case of plagiarism in literary translation. *The International Journal of Speech*, *Language and the Law*, 11(1), 1-26

_____. (2008), Plagiarism. In Gibbons, J., and Turell, M. (eds.) *Dimensions of Forensic Linguistics*. Philadelphia: John Benjamins Publishing Company.

Appendix

Seventeen Samples of Research Articles Matrix

Author	Year	Segment	The	of	and	in	to	A	is	that	it	as	this	are	not	be	for	but	with	at
A1	1967	1	83	41	21	30	49	28	30	17	18	18	10	6	14	12	8	4	2	3
A2		3	81	54	38	23	43	22	23	13	15	8	8	7	7	10	8	4	8	6
A3	1977	4	49	45	38	24	18	21	23	14	10	12	10	12	6	10	1	6	6	5
A4		5	52	37	34	26	22	29	34	17	14	4	16	9	14	2	7	6	12	5
A5		6	65	55	32	31	37	20	10	11	14	17	14	6	7	3	6	10	7	11
A6	1988	7	63	35	31	23	30	23	7	24	5	13	3	5	8	6	7	8	5	7
A7		8	67	61	37	35	25	15	11	25	9	8	11	6	2	8	6	3	10	7
A8	1998	10	50	35	38	36	23	18	31	18	19	9	13	15	14	4	8	8	10	4
A9		11	65	35	42	36	22	20	10	15	9	13	7	5	2	9	11	6	8	2
A10		12	54	30	28	23	30	17	22	18	15	8	11	16	5	14	10	8	5	8
A11	2002	13	57	41	27	24	31	19	16	23	13	16	14	2	7	7	2	8	3	6
Phili12	2010	16	71	62	40	33	21	15	7	3	4	6	5	10	1	1	12	0	1	1
Phili 13		17	68	64	36	38	18	16	3	6	0	6	5	2	5	0	15	2	2	0
Iraqi 14	2005	18	74	48	23	21	34	14	14	17	10	16	12	19	4	4	0	3	9	7
Iraqi 15		19	99	32	25	40	18	18	30	17	6	15	10	12	5	6	0	1	3	1
Mixed 16	2012	20	48	60	35	24	22	18	7	10	3	7	8	7	2	5	10	1	4	2
Mixed 17	2012	21	46	34	42	37	24	28	17	7	6	15	10	8	8	6	9	2	5	2

which	has	on	what	there	so	from	other	can	all
4	13	3	8	3	3	4	3	4	3
2	10	1	4	4	2	1	3	2	7
5	2	1	6	4	5	5	3	11	2
2	3	3	6	2	2	1	3	6	3
9	6	8	4	1	6	7	3	2	1
7	3	7	5	3	7	1	2	2	2
5	2	5	4	6	8	7	6	1	1
4	8	5	1	10	5	2	7	2	4
2	12	8	5	4	3	7	8	3	4
3	4	6	3	6	5	6	6	3	4
6	1	4	5	3	3	5	1	10	2
3	5	7	1	0	0	5	1	2	0
1	1	2	0	1	0	3	0	0	0
4	3	3	3	1	8	8	3	1	7
11	3	1	3	6	1	5	2	2	2
1	4	6	0	3	0	15	6	1	1
11	4	3	0	0	0	6	7	2	1