

# VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP

**Marwa Hassan Mohammed AlNagheya**

Assistant Lecturer, Department of English, Port Said University

mohammed.elsayed@arts.psu.edu.eg

## **Abstract**

In the age of the Internet and social media, where many writers hide behind their blue screens, broadcasting all kinds of ideas and committing new types of crimes through steering public opinion sheltered with the courage of impunity, the ability to use linguistic tools as evidence of either guilt or innocence becomes of utmost importance. The present study is set to investigate the potential linguistic potency of unmasking the hidden identity of prominent authors whose ideas get the chance to be published opposite the editorial page in leading national newspaper in Egypt under a pseudonym. Pseudonymous op-eds authorship is expected to be the most hidden among other genres of writing. This is because the text is expected, as a rule of thumb, to be reviewed and proofread by the publisher and editorial board before getting to appear in a newspaper. The study at hand centers around the hypothesis that there must be linguistic features that survive the alterations made to the text that can still lead to the real hand holding the pen behind the post-editorial pen-named column. The study employs a corpus-based approach to analyzing stylometric features in an article published in Al-Masry Al-Youm Newspaper's column وجدتها *wajadtoha* [Eureka] that is regularly signed *Newton* for the end goal of getting its authorship verified. A contrastive analysis was conducted against ArabiCorpus — an online corpus of Brigham Young University (BYU). Using the stylometric features detected in the pen-named op-ed as ruling criteria, the finding revealed a notable variance from the way they exhibit on the sub corpus [all newspaper] of the online ArabiCorpus revealing distinctiveness. The analysis presented in this paper forms the first stage of any

possible authorship dispute be it verification or attribution through the creation of a forensic linguistic profile of the disputed text. This way it is expected to contribute to the field of forensic linguistics by demonstrating the efficacy of stylometric tools in resolving authorship disputes in two challenging scopes: the genre of writing like op-eds and the less researched Arabic language.

**Keywords:** pseudonym, op-ed, Al-Masry Al-Youm, Newton's Eureka, Egyptian-Arabic newspaper, stylometry, authorship attribution, forensic linguistics, corpus linguistics

### introduction

In the realm of journalism, often referred to as the fourth estate, op-eds hold a significant position by which they can steer public opinion. Op-eds are concise opinion articles that get to be published opposite the editorial page, hence their significance. They are typically the product of an independent voice that is not part of the editorial team. When these columns try to advocate a critical point of view and also have the identity of its author blocked behind a pseudonym, a forensic concern emerges.

On April 17, 2020, during a live Egyptian talk show, it was disclosed that the Supreme Council for Media Regulation (SCMR) had summoned the author of the *Al-Masry Al-Youm* newspaper column *وجدتها Wajadtoha* [Eureka]—a column regularly signed under the pseudonym "Newton"—for investigation. Furthermore, it was announced that the publication of all articles authored by this individual would be suspended for one month. This action followed the publication of an article titled *استحداث وظيفة qistihdāth wazīfa* [The Introduction of a New Job], in which the columnist advocated for granting administrative autonomy to the Sinai Peninsula and called for the appointment of an independent "governor" to oversee the region (Al-Hekaya, 2020). This article provoked harsh public criticism. A rather exasperated party was Khaled Miri, editor-in-chief of *Al-Akhbar* newspaper, who expressed disapproval in his April 14 column titled *صفقات مشبوهة safqāt mashbūha* [Suspicious Deals]. In this article, Miri skeptically questioned both Newton's agenda, political loyalties,

**VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A  
STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP**

Marwa Hassan Mohammed AlNagheya

---

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

and above all the tolerance of using pseudonyms in Egyptian journalism. He also suggested suspicious ulterior motives behind the columnist's proposal and urged the Supreme Council for Media Regulation (SCMR) to disclose the columnist's identity, citizenship, and affiliation (Miri, 2020). The previous scenario presents itself then as a forensic situation where an authorship needs to be verified even if someone later came forward and claimed it or a scapegoat was blamed for it.

The question "Does the field of linguistics possess sufficiently robust forensic evidence to either substantiate or refute claims of authoring any written text, in a manner comparable to how criminal investigation bodies use fingerprint or DNA analysis?" posed itself. Nevertheless, a sample text known to be written by a suspect author remains needed to make matching possible. A tip that satisfied this need came on the tongue of the talk show host, Amr Adib, as he was commenting on this controversy. Although Diab noted that the true authorship of this disputed article remained uncertain, he mentioned that the Newton column is rumored to be most likely written by Salah Diab, founder and owner of the Al-Masry Al-Youm Newspaper. In a live telephone interview during the same episode with Al-Masry Al-Youm's editor-in-chief, Abdel Latif El-Menawy, Adib directly asked who had authored the column in question. El-Menawy responded with a pause and a smile, suggesting that the author might indeed be someone previously mentioned by Adib during the broadcast.

### **Literature Review**

The proliferation of pseudonymous texts in digital platforms in all their forms gave rise to an intense need for a linguistically grounded method to verify their authorship. As a matter of course, authors need to bear reliability of their words particularly when they yield a legal situation. Literature review in the present study is structured as presenting key concepts to the study, standing on the latest approaches to authorship studies in the past years and

whether the genre of opinion articles was put to linguistic analysis for this end before.

Authorship profiling is an endeavor that busied many researchers in various driplines like linguistics, digital text forensics (e.g., copyright cases), and computational engineering. Generally speaking, it is the detection of unconscious choices an author made as he/she produces any text. This way it aligns with the trend that advocates there is a "textual/linguistic fingerprint" (Baayen et al., 2002; Berges, 2015; Elek, 2016) that can single out the author especially when "not consciously changing their style".

an answer to a yes/no question of authorship (Boenninghoff et al., 2019; Halvani et al., 2016; Zhang et al., 2021, Stamatatos, 2025). In its simplest context, it runs a comparison between two texts, one with a known author and the other's is disputed, with the intention of arriving at a conclusion to whether they are produced by the same author or not. Although the term "authorship attribution" may be more recurrent in research, it is of vital importance to put effort in researching verification for it is as true as Stamatatos put it in 2016 when he mentioned that any case of authorship attribution is the sum of a number of verification problems (Stamatatos, 2016). Either is expected to follow linguistic traces of "an unconscious personal stamp" (Eder, 2011) of whoever is responsible for the production of the text.

Stylometricians had initially attested to the claim that style can be as certain to idiosyncratically identify its creator as handwriting can (e.g, Lutosławski (1897: 66; Michaelson et al., 1979) and even be fit for a "biometric identification (Chaski, 2001). More recent research sees that no features can remain invariable across genre and time (Nini, 2023). Some features have passed the time-genre filter without any inconsistency since Mosteller and Wallace attribution of the Federalist Papers in 1963. The first of these features is the frequency of function words – functors- like prepositions, pronouns and articles. They were determined to be key to any analysis of authorship by Kestemont (2014). This sounds sensible because function words are not charged with genre-specific meanings. Functors also transcend the time criterion

## VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP

Marwa Hassan Mohammed AlNagheya

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

and kept discriminating authors thus far (Argamon, 2018; Argamon & Levitan, 2005; Kestemont, 2014; Grieve, 2007 & Stamatatos, 2009).

Nini added a critical shift to this conclusion to which computer science studies had relied on for decades. Nini demonstrated that the use of function words is not in itself an authorship criterion in terms of individual preference; it is rather a “a readable signs of how grammar is being realized and used. He illustrated it further by showing that it is never the tendency to use a word like *that* for example that can discriminate an author from another, but how this word functions in its context a *demonstrative pronoun*, a *relative pronoun*, or a *complementizer* for instance (Nini, 2023). This idea relates to Carole Chaski’s position on the reliability of syntax in arriving at an authorial writer’s print or a *fingerprint* in Chaski’s terms. She expressed how while authors get consumed in producing meaning, their grammar remains unconscious and thus hard to fake (Chaski, 1997). To her, even punctuation marks are analyzed in relation to their syntactic role not in the selection of a mark over another. Although her analysis method is purely computational, she entirely rejects computed calculations of things like the length of words or sentences and judges it as atheoretical and outside the scientific arena. The frequency of n-grams is the second of those time-genre-surpassing features.

### Data and Method

The study at hand puts the disputed op-ed استحداث وظيفة *aiṣṭiḥdāth wazīfa* [The Introduction of a New Job] to analysis with the aim of collecting the stylometric features that were not obliterated by the work of the newspaper editor, i.e., giving character to the writing of this very particular unknown author. Given the fact that this pen-named article gave rise to a terrible legal ordeal to both the author and the host newspaper resulting in a definitive ban on publishing under a pseudonym in the Egyptian press without prior written approval from the Supreme Council for Media Regulation (SCMR), a forensic situation hereupon forms.

An infuriated official journalist reacted to the opinions expressed in the article by filing a complaint to the SCMR against the newspaper that was consequently fined with a 250 thousand Egyptian pounds. Following this, the yet-so-far anonymous author earned an absolute suspension of publication and was later referred to Prosecutor General. Owing to the escalations, the disputed article was removed from the website of *Al-Masry Al-Youm* in submission to an official decree by the SCMR; therefore, it had been collected from the print edition of the issue number ٥٧٨١ dated 12/4/2020 that has an online access at <https://www.almasryalyoum.com/news/details/2220481>.

The methodological foundation guiding the analysis, then, rests on three levels. The first is a qualitative stylometric analysis of the op-ed. In this stage, the article is analyzed for a) structural features, b) syntactic complexity, b) lexical features, and d) salient stylometric features. The second level eclectically employs the framework used by Nini (2015) to help create a profile of the unknown author linking him or her to a social group that is characterized by age, gender, and social class. The final stage is entirely quantitative for gaining evidence on how distinctive and unique the stylometric profile is given the fact that the text is published in an edited and proofread newspaper and might have underwent alterations. In this level, the elicited linguistic patterns are to be measured against the online Arabi Corpus where [Al-Masry Al-Youm] and [All Newspapers] are sub-genres. ArabiCorpus is a free untagged Brigham Young University 30-million-word corpus). COCA (Corpus of Contemporary American English) is also visited to compare lexical selection frequencies to reach better results in profiling cultural background.

### **Forensic Authorship Analysis**

The analysis herein is set to collect all salient features of the way language is used by the disputed author in a pursuit of retrieving a measurable pattern that is unique to this very author. Although the features can be endless, the present study tries to

**VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A  
STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP**

Marwa Hassan Mohammed AlNagheya

---

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

limit the analysis to the following categories: structural, lexical, syntactic.

**A. *Structural Features***

In this section, the organization and formatting of the text are described. The text in the op-ed is organized in a rather concise way with no headings, subheadings, bullet points, or any formatting features like italic/bold text, not even an indentation (although the paragraphs are indented in the printed edition). This would not allow analysis to rely much on these features to attribute a distinct signature in terms of text organization, although absence of features is a feature in itself. The author paragraphed his/her article into 12 very short paragraphs. The longest is made up of only four lines (58 words), and the shortest is single lined (3 words). The average paragraph length is two lines. The commencing paragraph is an introduction that paves the way to the striking proposition to come (48 words). The following paragraph (27 words) names the job that the article title excites the reader about. For 9 coming paragraphs, the author discusses the job description (31 words), how requirements of the job can be met (23 words), powers of the job holder (17 words), supporting experiences to the proposed idea (46 words), raises a question about the possible consequences of his proposition (2 words), and then presents positive opportunities (25 words), support them (29 words), and then ends with citing how negative results will not even trouble anybody. The conclusion is formed with an open-ended rhetorical question that reads in English as “So, I wonder, is it possible to have this experiment undertaken and see Sinai treated as a special economic zone? Or are the obstacles to this experiment too great to overcome?” (25 words). The heaviest weight of words (58) is given to the paragraph in which the author takes a stance against whoever might be against the idea.

**B. *Lexical Features***

Regarding parts of speech (POS), The article is rich in nouns. They make up roughly half of the text (i.e., 49.12%). The next higher frequency of POS in this article is found to be prepositions

(i.e., 16.66%). Pronouns came third in frequency (14.32%). Out of the almost-50% of nouns, proper names make 16%. Examples are دينج شياو بينج *dynj shyaw bynj* [Deng Xiaoping], مهاتير محمد *mahātīru muḥamād* [Mahathir Mohamad], الصين الحديثة *qīlṣīan alḥadythā* [Modern China], ماليزيا *mālīzīā* [Malaysia] (×2), سيناء *saynā* [Sinai] (×4), سنغافورة *siṅghāfūra* [Singapore] (×2), هونغ كونج *hūṅg kūṅg* [Hong Kong] (×2), الأردن *alo'rdn* [Jordan], كوريا الجنوبية *kūrīā aljanūbīā* [South Korea], جمال حمدان *gamāl ḥamdān* [Gamal Hamdan], and عبقرية الزمان والمكان *'abqarīātū alzāmāni wālmakāni* [Genius of Time and Place].

Lexical density was calculated through dividing the total of lexical words in the text (160) by the sum of all the words in it (342). This equation resulted in a lexical density of 0.467 that is 46.78% of the whole disputed text.

For lexical richness, a different measure is taken. This time, it is not a certain POS that is taken into consideration, but rather words that have a “unique” encounter. Type-token ratio (TTR) measure is used to inform the analysis on how far the author’s lexical repertoire can be judged as diverse and rich. It is calculated using the formula presented in Figure 3 through dividing the number of unique words (types) by the number of the total words in the article (tokens). The lower TTR gets, the more repetitions of lexical items can be detected leading to a lower lexical diversity and richness are there in the author’s writing repertoire. This resulted in the following realization: (262 TYP ÷ 343 TOK) × 100 = 76.37% TTR.

### Figure 3

Type-Token Ratio Formula

$$TTR = \left( \frac{TYP}{TOK} \right) \times 100$$

Numbers of both types and tokens had been retrieved through AntConc software version 4.2.4. (Anthony, 2023) see Figure 4.



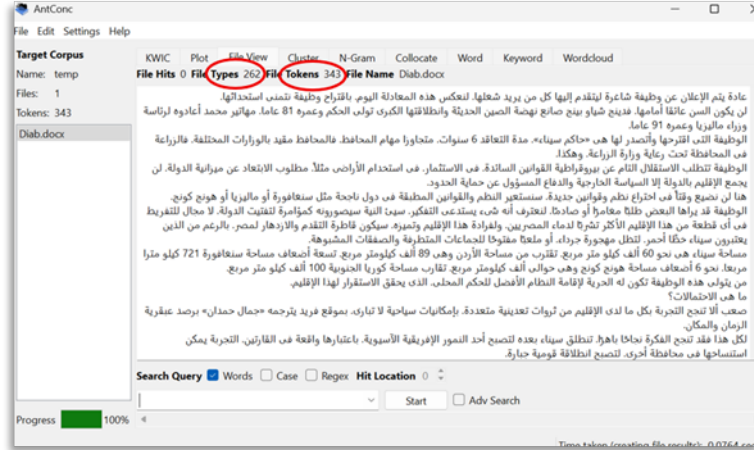
## VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP

Marwa Hassan Mohammed AlNagheya

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

Figure 4

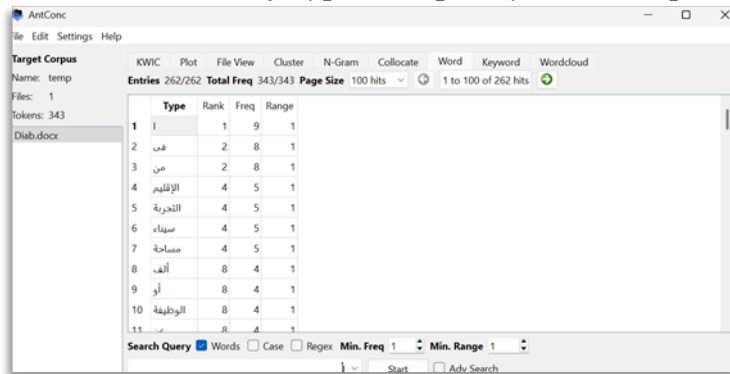
Results of AntConc Calculation of TTR



The top frequently recurrent word in this text is found to be not only one but two in a draw في /fī/ (in) and عن /ʿan/ (about; from) with an equal frequency of 8 times at 2.33%. في stands out as used markedly for the author keeps typing it throughout the article using the keyboard strokes (T + N) although even Microsoft Office Auto corrector marks it with a red squiggly line signaling a potential spelling error.

Figure 5

AntConc Wordlist of Types Frequency in the Disputed Op-Ed



### C. Syntactic Features

The op-ed article is intended to be examined for the features that can conjointly create a traceable linguistic print of the author.

**Syntactic Complexity:** The op-ed, *The Introduction of a New Job*, is made up of 342 words that are segmented into 48 sentences, given the end mark criteria as put by the author. This segmentation does not always correspond to the grammatically accepted structure of a sentence and results in various fragments (e.g., *in investment, in land use, and bypassing the duties of a governor*). However, it is not about accuracy here; it is all about how he/she puts things together in a way that can discriminate their writing. Consequently, the component words were grouped into syntactic units (see Table 1). To examine the syntactic complexity of the sentences combining the article, two types of measures are taken into count: clause-based measures and structural measures. For clause-based measures, ratio of clauses to sentences, subordinate clause ratio, and coordination index are calculated. To better arrive at the clauses, the text was first translated to English and grouped into T-units. A single T-unit is defined by Kellogg Hunt as “one main clause together with all attached subordinate clauses” (1965, p.20). Hunt’s definition would break a compound sentence, then, to two T-units and in this study counts for the syntactic pattern subject-verb-object/object complement. The text is segmented into the T-units it makes to allow for measuring their mean length. This is done by dividing the total number of the op-ed articles words (342) by the total number of t-units (27). This results in an average of 12.66 words per T-unit. The 27 T-units in the article have a mean number of clauses per each. (see Table 1).

**VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A  
STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP**

Marwa Hassan Mohammed AlNagheya

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

**Table 1**

*An English Translation of the Disputed Op-Ed Broken into  
Syntactic Units*

No.	Sentence	Clause	Clause Type	T-unit
1	عادة يتم الإعلان عن وظيفة شاغرة ليتقدم إليها كل من يريد شغلها. Usually, a vacancy is announced for anyone who wants to apply.	<i>Usually, a vacancy is announced for anyone...</i> <i>who wants to apply</i>	main relative	1
2	لنعكس هذه المعادلة اليوم. باقتراح وظيفة نتمنى استحداثها Let's reverse this equation today by proposing a job we hope will be introduced.	<i>Let's reverse this equation ...</i> <i>we hope [that]...</i> <i>...will be introduced.</i>	main reduced relative embedded object	1
3	لن يكون السن عائقا أمامها Age will not be a barrier to it. فدينج شياو بينج صانع نهضة الصين الحديثة وانطلاقتها الكبرى تولى الحكم وعمره ٨١ عاما As Deng Xiaoping, the architect of China's modern renaissance and its great breakthrough, assumed power at the age of 81. مهاتير محمد أعادوه لرئاسة وزراء ماليزيا وعمره ٩١ عاما Mahathir Mohamad was reinstated in the Prime Ministry of Malaysia at the age of 91.	<i>Age will not be a barrier to it.</i> <i>As Deng Xiaoping ...</i>	main adverbial causal	1
4	الوظيفة التي اقترحها وأتصدر لها هي «حاكم سيناء». The job I am proposing and applying for is «Governor of Sinai».	<i>The job .... is</i> <i>I am proposing ....</i>	main reduced relative	1
5	مدة التعاقد ٦ سنوات The contract is for a six-year term. متجاوزا مهام المحافظ. bypassing the duties of a governor, فالمحافظ مقيد بالوزارات المختلفة. since the governor is	<i>The contract is for ...</i> <i>bypassing the duties of a governor.</i>	main embedded non-finite present participle clause	1

	bound by various ministries, فالزراعة في المحافظة تحت رعاية وزارة الزراعة. وهكذا inasmuch as agriculture in the governorate falls under the auspices of the Ministry of Agriculture, and so on.	since the governor is bound by various ministries.	subordinate causal	
		inasmuch as agriculture in the governorate falls under ...	subordinate causal	
6	الوظيفة تتطلب الاستقلال التام عن بيروقراطية القوانين السائدة. The position requires complete independence from the bureaucracy of prevailing laws. في الاستثمار. In investment. في استخدام الأراضي مثلاً. In land use, for example.	The position requires...	main	1
7	مطلوب الابتعاد عن ميزانية الدولة. It is required to get distanced from the state budget.	It is required ...	main	1
		to get distanced from the state budget.	infinitival clause acting as extraposed subject	
8	لن يجمع الإقليم بالدولة إلا السياسة الخارجية والدفاع المسؤول عن حماية الحدود. The only things that will link the region to the state are foreign policy and defense, which is responsible for protecting the borders.	The only things ...are foreign policy and defense	main	1
		that will link the region to the state	relative adjective clause	
		which is responsible for protecting the borders.	relative adjective clause	
9	هنا لن نضيع وقتاً في اختراع نظم وقوانين جديدة. Here, we will not waste time inventing new systems and laws.	we will not waste time inventing new systems and laws.	main finite	1
		inventing new systems and laws.	non-finite complement	
10	سنستعير النظم والقوانين المطبقة في دول ناجحة مثل سنغافورة أو ماليزيا أو هونغ كونج. We will borrow the systems and laws applied in successful countries like Singapore, Malaysia, and Hong	We will borrow the systems and laws ... applied in successful countries like Singapore, Malaysia, and Hong Kong.	main finite	1
			reduced relative	

**VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A  
STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP**

Marwa Hassan Mohammed AlNagheya

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

	Kong.			
11	الوظيفة قد يراها البعض طلبًا مغامرًا أو صادمًا. Some may see the position as an adventurous or shocking request.	<i>Some may see the position as an adventurous or shocking request.</i>	main finite	1
12	لنعترف أنه شيء يستدعي التفكير. Let's admit it's something to think about.	<i>Let's ...</i>	imperative main clause	1
		<i>it's something to think about.</i>	embedded reduced relative	
13	سيئ النية سيصورونه كمؤامرة لتفكيك الدولة. Those with ill intentions will portray it as a conspiracy to dismantle the state.	<i>Those with ill intentions will ...</i>	main	1
		<i>to dismantle the state</i>	embedded infinitival complement clause	
14	لا مجال للتقريط في أى قطعة من هذا الإقليم الأكثر تشربًا لدماء المصريين. There's no room for compromising any piece of this region which is the most steeped in the blood of Egyptians.	<i>There's no room ...</i>	main	1
		<i>which is the most steeped in the blood of Egyptians.</i>	relative clause	
15	ولفرادة هذا الإقليم وتميزه And given its uniqueness and distinctiveness. سيكون قاطرة التقدم والازدهار لمصر. It will be the engine of progress and prosperity for Egypt. بالرغم من الذين يعتبرون سيناء خطأ أحمر. Despite those who consider Sinai a red line. لتظل مهجورة جرداء To remain deserted and barren. أو ملعبًا مفتوحًا للجماعات المتطرفة والصفقات المشبوهة Or an open playground for extremist groups and dubious deals.	<i>It will be ....</i>	main clause	1
		<i>who consider Sinai a red line</i>	relative clause	
		<i>to remain deserted and barren</i>	embedded infinitival complement clause	

16	مساحة سيناء هي نحو ٦٠ ألف كيلو متر مربع. The Sinai Peninsula covers an area of approximately 60 thousand square kilometers.		main clause	1
17	تقترب من مساحة الأردن وهي ٨٩ ألف كيلومتر مربع. It's close to the size of Jordan, which is 89 thousand square kilometers. تسعة أضعاف مساحة سنغافورة ٧٢١ كيلو مترا مربعا. Nine times the size of Singapore 721 square kilometers. نحو ٦ أضعاف مساحة هونغ كونج وهي حوالي ألف كيلومتر مربع. nearly 6 times the size of Hong Kong, at about thousand square kilometers. تقارب مساحة كوريا الجنوبية ١٠٠ ألف كيلومتر مربع. Close to the size of South Korea 100 thousand square kilometers.	<i>It's close to the size of Jordan</i>	main	1
		<i>which is 89,000 square kilometers.</i>	non-restrictive relative clause	
		<i>Close to the size of South Korea</i>	reduced clause	
18	من يتولى هذه الوظيفة تكون له الحرية لإقامة النظام الأفضل للحكم المحلي. Whoever takes over this job shall have the freedom to establish the best system of local governance. الذي يحقق الاستقرار لهذا الإقليم. that achieves stability for this region.	<i>Whoever assumes this position</i>	nominal	1
		<i>that achieves stability for this region.</i>		
		<i>... will have the freedom</i>	main clause	
		<i>to establish the best system</i>	infinitival clause	
19	ما هي الاحتمالات؟ What are the chances/odds?	<i>that achieves stability for this region.</i>	relative clause	1
			main clause	

**VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A  
STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP**

Marwa Hassan Mohammed AlNagheya

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

20	صعب ألا تنجح التجربة بكل ما لدى الإقليم من ثروات تعدينية متعددة It's hard for the experiment not to succeed, given the region's diverse mineral wealth.  بإمكانات سياحية لا تبارى. with unparalleled tourism potential.  بموقع فريد يترجمه «جمال حمدان» برصد عبقرية الزمان والمكان With a unique location, translated by «Gamal Hamdan» with a capture of the genius of time and place.	<i>It's hard</i>	cleft anticipatory it main clause	1
		<i>for the experiment not to succeed</i>	delayed complement infinitival subject clause	
		<i>translated by Gamal Hamdan with a capture of the genius of time and place.</i>	reduced adjective clause	
21	لكل هذا فقد تنجح الفكرة نجاحاً باهراً. For all this, the idea may succeed resoundingly.	<i>...the idea may succeed resoundingly.</i>	main finite clause	1
22	تنطلق سيناء بعده لتصبح أحد النمور الإفريقية الآسيوية. Sinai will then take off to become one of the African- Asian tigers.  باعتبارها واقعة في القارتين. Given the fact that it's located on both continents.	<i>Sinai will then take off...</i>	main	1
		<i>to become one of the African-Asian tigers.</i>	non-finite infinitival clause	
		<i>Given the fact that...</i>	non-finite participial clause functioning as an adverbial	
		<i>...it's located on both continents.</i>	embedded finite noun clause, acting as the complement of the fact	

23	<p>التجربة يمكن استنساخها في محافظة أخرى. The experiment could be cloned in another governorate.</p> <p>لتصبح انطلاقة قومية جبارة. So it becomes a massive national breakthrough.</p>	<i>The experiment could be cloned in another governorate.</i>	finite independent clause acting as a cause	1
		So it becomes a massive national breakthrough.	coordinated finite independent clause acting as an effect	
24	<p>أما إذا فشلت التجربة. But if the experiment fails.</p> <p>فليس هناك الكثير لكى نبكى عليه. there's not much to cry over, then.</p>	<i>if the experiment fails.</i>	finite conditional subordinating clause	1
		<i>there's not much to cry over.</i>	finite independent clause	
25	<p>فهل يا ترى من الوارد القيام بهذه التجربة والتعامل مع سيناء كمنطقة اقتصادية ذات طبيعة خاصة. So, I wonder, is it possible to have this experiment undertaken and see Sinai treated as a special economic zone?</p> <p>أم أن معوقات هذه التجربة أكبر من أن نتجاوزها؟ Or are the obstacles to this experiment too great to overcome?</p>	So, I wonder.	main clause	3
		<i>Is it possible</i>	embedded interrogative clause	
		<i>to have this experiment undertaken...</i>	non-finite infinitive clause	
		<i>and see Sinai treated as a special economic zone?</i>	coordinated non-finite infinitive clause	
		<i>Or are the obstacles to this experiment too great to overcome?</i>	coordinated interrogative clause	
Total	25	59		27

#### D. Salient Markers

The use of language in this text revealed several distinguishing marks that can add to the linguistic preferences of the disputed author creating a profile. Since the text is in Arabic, this allowed the analysis the chance to examine things like punctuation marks, use of diacritics, keyboard key choices, lexical choices, digit/letter numbers, spacing and number/gender agreement.



**Punctuation.** Use of punctuation marks shows salient choices. For quotes, the author resides to using pointed double angle quotation marks, *guillemets*, to enclose information (e.g. « حاكم «جمال حمدان» », «سيناء»). Other than this variety of quotation marks and a thrice use of question marks, no punctuation mark is used but the full stop, and it is not used to end a well-formed sentence but rather ideas (e.g., «متجاوزا مهام المحافظ. باقتراح وظيفة نتمنى استحداثها. في استخدام (لتظل مهجورة جرداء. فالمحافظ مقيد بالوزارات المختلفة. الأراضى مثلاً»). The author pauses with a full stop no matter what grammar rules allow.

**Orthography.** The way words are typed (i.e., keyboard selections) also has something to say here. Initial glottal stops with FatHah [macron above] are written with the Hamza /ء/ added above the letter Alif diacritically 15 times out of the 16 cases in the article (93.7%). Initial glottal stops with Kasra [underbar below] are written with the diacritic Hamza below the letter Alif only 11 times out of the 25 cases in the article (44%). The flaccid Alif with the long vowel, *madda*, is used once in *الآسيوية* *alāsiwīyātū* [Asian]. The author places two dots over the knotted Taa' 59 times, setting it apart from Haa' (using two different keyboard strokes). Out of the 11 nunated nouns, nine received Tanween [nunation diacritic] above the penultimate letter 87.5% (e.g., «نجاحًا باهرًا»). Five words were not nunated at all where they should have been (e.g., «علماء»، «عائقًا»). Only two times was the nunation given to the ultimate letter (i.e., «وقتًا»، «مثلاً»). Another notable usage is the way the letter ي is typed at the word end boundary. The author appears not to differentiate between ى which is typed by pressing the N button on the keyboard and ي that is made by the D button. All 15 occurrences of final position ي are produced with the N-button variety even where ي is needed (e.g., «فى»، «الأراضى»، «هى»، «حوالى»، «أى»، «الذى»). None of them receives two under dots. A 100% of the times the word *شيء* *shay'ūn* [thing] is used, follow this variety of keyboard spelling شىء using three keyboard keys (A, N, X), not شيء that requires striking the (A-D-Z) buttons. Another 100% is obtained in the analysis with the word *سيء* *sī'a* [bad] as it is typed using two keyboard strokes (S and Z) each time it is mentioned.

Two cases of disagreement with gender are detected ( لتصبح سيناء أحد ) (سبئ النية سيصورونه كمؤامرة & النمر).

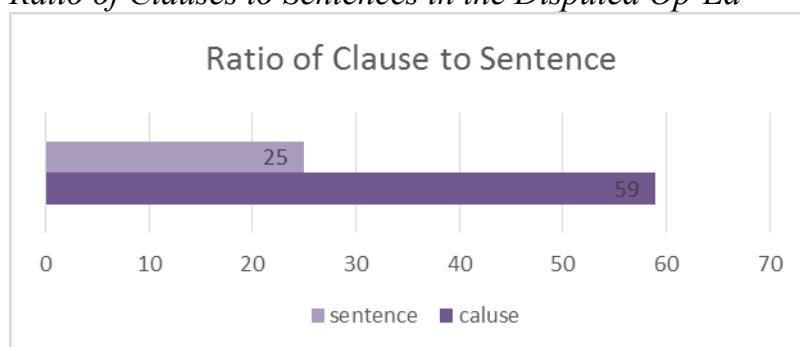
**Numbers.** They are almost entirely given in digits except for the word ألف [thousand], is always spelled out (e.g., 81 عاما، ٩١ عاما، ٨٩ ألف كيلو متر مربع، ٧٢١ كيلو ٦ سنوات، ٦ أضعاف، ٦٠ ألف كيلو متر مربع، ٨٩ ألف كيلو متر مربع، ١٠٠ ألف كيلو متر مربع (تسعة أضعاف). The author used the letter variety in giving number only one time (تسعة أضعاف). It was not just spelled out, but moreover grammatically correct in terms of tamyiz. He used an indefinite genitive plural noun أضعاف with a number that has a reversed gender agreement تسعة. He used no tanwīn nunation diacritics at all in delivering numbers.

### Results

The article was found to heavily employ clauses with an average of 59 clauses that exceed the 25 sentences by more than the double (see Figure 6). Of these clauses, only 3 are in subordination (*since the governor is bound by various ministries...*, *inasmuch as agriculture ... falls under...*, and *if the experiment fails*), and 2 are coordinated (*Or are the obstacles to this ... too great to overcome*, and *see Sinai treated as a special economic zone*). Most of the clauses used to present the author's mind are embedded (e.g., *So, I wonder, is it possible to have this experiment undertaken*, *Given the fact that it's located on both continents*, *Despite those who consider Sinai a red line to remain deserted and barren*, *those who with ill intentions will portray it as ... to dismantle the state*, *Let's admit it's something to think about*).

**Figure 6**

*Ratio of Clauses to Sentences in the Disputed Op-Ed*



# VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP

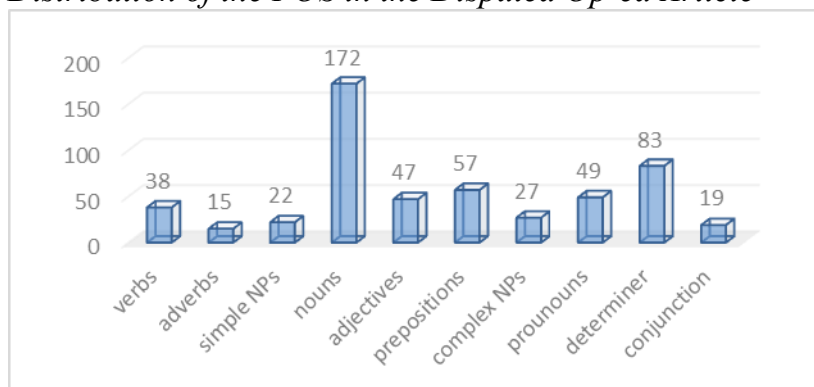
Marwa Hassan Mohammed AlNagheya

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

Despite the common syntactic disconnection and breakage of sentences into smaller units of either phrases or clauses lacking a main clause within the boundaries of the sentence, the author's text is found to mostly follow a **nominal style**. Nominal style markers (i.e., nouns, adjectives, prepositions, and complex noun phrases) are found to form 72.23% of the article. Proper names and adjectives are highly employed by the author of the disputed article to give support to his proposition with real examples of the experience he is presenting. This is clear from the frequent use of proper nouns and proper adjectives in his article (e.g. دينج شياو بينج، مهاثير محمد، الصين الحديثة، ماليزيا، سيناء، سنغافورة، ماليزيا، هونج كونج، الأردن، (سنغافورة، هونج كونج، كوريا الجنوبية، جمال حمدان، عبقرية الزمان والمكان). On the flip side, **clausal style** that is identified by use of verbs, adverbs, and simple noun phrases that comprise only a pronoun is also tallied and calculated as 21.8 % of the total word count of the article. POS count as used to present propositions in the disputed article is presented in Figure 2.

**Figure 7**

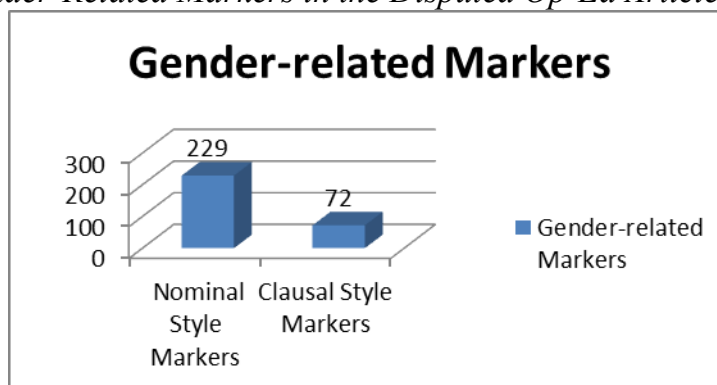
*Distribution of the POS in the Disputed Op-ed Article*



Predominance of nominal style over clausal style in the questioned article's author (see Figure 8) allows for the belief that it is more likely male. This result aligns with what Nini (2018) arrived at from the studies he reviewed to apply their results in an experimental context in his study.

**Figure 8**

*Gender-Related Markers in the Disputed Op-Ed Article*



According to Nini's empirical framework, the author here is not young of age, linguistically. This can be evidenced by a few observations. First, ageing has been linked to the use of nominal style already observed for gender (see Figure 8). Another reason is that the author makes use of a higher frequency of determiners and prepositions than that of pronouns (see Figure 7). A third reason is the high type-token ratio of the text (76.37% TTR) and the high lexical diversity and richness that comes with it. This aligns with the results of empirical studies prior to the present one which had found that with ageing, there comes a rise in using determiners and prepositions and fall in using pronouns. (Nini, 2018; Pennebaker and Stone, 2003). A good reason why the author is thought to be of higher age is his use of "informative style". Upon proposing an idea in his article, the author of the disputed article is seen to be educating the readers and enriching their knowledge with real examples of the experience he is presenting. This is clear from the frequent use of proper nouns and proper adjectives in his article (e.g. دينج شياو بينج، مهاتير محمد، الصين الحديثة، ماليزيا، سيناء، سنغافورة، ماليزيا، هونج كونج، الأردن، سنغافورة، هونج كونج، كوريا الجنوبية، جمال حمدان، عبقرية (الزمان والمكان).

Lexical density was calculated through dividing the total of lexical words in the text (160) by the sum of all the words in it (342). This equation resulted in a lexical density of 0.467, which is 46.78% of the whole disputed text. Manual Berman (2008) Higher education.

Stemming from the empirical hypothesis that language is recycled (Nini, 2023) and mostly prefabricated (Erman & Wren, 2000), the present study believes that the author at hand recycles English constructions to Arabic output. These lexical choices stand out as contributive clues to his cultural and social background. The writer is thought to be inclined to the American culture and language. This can be elicited from his lexical selection of terms like *حاكم سيناء* *ḥākīmūn sayunā'* [governor of Sinai], *الإقليم* *al-ʿaqlīm* [the region], *ما هي الاحتمالات؟* *mā hī al-aḥtmālāt* [what are the odds], and *الإفريقية الآسيوية* *al-ʾafriqīyā al-āsiyāwya* [African-Asian]. This is deduced from a number of lexical choices. The word *حاكم*, governor in English, that he chose to name the new job introduced «حاكم» relates to a position in the American federal system where each of the fifty states has an independent governor as its head. The author suggested not only the Sinai has its own governor "independent of the prevailing bureaucracy", but also to "clone the experience to another governorate" and this relates heavily to the American reference. Another American-background-telling choice is the proper adjective *الإفريقية الآسيوية*, [African Asian]. The adjective is looked up in the ArabiCorpus (a free 30-million-word corpus provided by Brigham Young University found at <http://arabicorpus.byu.edu>). The corpus search resulted in that *الإفريقية الآسيوية* has a total frequency of five occurrences on the All-Newspapers sub-corpus. On the other hand, the other variety of the adjective (i.e., *الأفروآسيوية*) was found to be more frequent with a total frequency of 45 occurrences. This result was checked against the NOW Corpus (News On the Web) to find which variety is used in the same register (i.e., newspapers). The search was limited to collect data only from the United States section. The returned results showed that African Asian (i.e. the English equivalent of *الإفريقية الآسيوية*) is more frequent than Afroasiatic/Afroasian (i.e. the English equivalent to *أفروآسيوية*). This translates into the fact that the selection of *الإفريقية الآسيوية* relates to a tendency towards an American disposition (see Figure 9). A rather striking observation here is that the way this author types the term *الإفريقية الآسيوية* with

Initial glottal stops with *Kasrah* (underbar below) turns an even rarer use (only two times) with zero instances per 100.000 words in All newspapers; the thing that can lead us to the believe it is a usage unique to the disputed author.

**Figure 9**

*ArabiCorpus Results that Informs the Rare Use of the Term*

الأفريقية الآسيوية

summary of search results
CorpusName =News
word: AllfryqyQAlMsywyQ
search string: الأفريقية الآسيوية — AllfryqyQAlMsywyQ
database: All Newspapers
search time: 38 seconds
part of speech: adj
search part of speech: adj
total number of occurrences: 5
0 instances of الأفريقية الآسيوية per 100,000 words in All Newspapers.

summary of search results
CorpusName =News
word: AllfrwMsywyQ
search string: الأفروآسيوية — AllfrwMsywyQ
database: All Newspapers
search time: 29 seconds
part of speech: adj
search part of speech: adj
total number of occurrences: 45
0.03 instances of الأفروآسيوية per 100,000 words in All Newspapers.

A third lexical choice is the word إقليم *aq̣līm* [region]. The word was repeated four times in the article. Besides the idea of having regions in Egypt is not as common as it is in the United States, Sinai in the Egyptian culture is not considered a region unless combined with all three cities overlooking the Suez Canal (i.e. Port Said, Ismailia, and Suez). On COCA (Corpus of Contemporary American English), Sinai collocates in the Newspaper sub-corpus with adjectives like (Egyptian, Northern, Israeli, lawless, Southern, new, Christian, Golan, global, Islamist, militant, Jewish, southwestern, sacred, restive, neglected). These adjective-collocates give us a notion how this peninsula is being perceived in the American culture as indexed in their newspapers. It is not just a peninsula; it is a region ripped apart among different contrasting sides. Its identity is divided among Egypt (Egyptian), Israel (Israeli), and Syria (Golan). It is an exhibition of contrastive religious backgrounds (Christian, Islamist, and Jewish). It is also defined geographically (e.g. Northern, Southern, and Southwestern). Moreover, Sinai is attributed a negative semantic prosody with collocates such as (restive, lawless, critical, and neglected). So, selection of a terminology like ((إقليم) combined with a demonstrative pronoun (هذا) adds psychological distance to this

# VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP

Marwa Hassan Mohammed AlNagheya

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

part that usually, in contrast to this perception, collocates in Egyptian Newspapers with الحبيبة beloved. Occurrence of noun phrase as part of a prepositional phrase that is the object of the action expressed in (الاستقرار, stability in English) relates to the American Newspapers negative collocates (restive and lawless).

One more feature that links the author to the American culture is the use of the expression ما هي الاحتمالات؟. This string of words has zero occurrence on the Al-Masry Al-Youm sub-corpus on the ArabiCorpus with both diacritical varieties (see Figure 10). On the other hand, the English equivalent (what are the odds) is common on the same register on COCA.

**Figure 10**

*Frequency of the string ما هي الاحتمالات on Al-Masry Al-Youm Sub-Corpus on the BYU ArabiCorpus*

search results for mA hy AIAHtmAlAt   ما هي الاحتمالات in Al-Masri Al-Yawm 2010 download citations	search results for mA hy AIEHtmAlAt   ما هي الإحتمالات in Al-Masri Al-Yawm 2010 download citations
summary of search results	summary of search results
CorpusName =Masri2010 word: mA hy AIAHtmAlAt search string: — ما هي الاحتمالات — mA hy AIAHtmAlAt database: Al-Masri Al-Yawm 2010 search time: 2 seconds part of speech: string search part of speech: string total number of occurrences: 0  0 instances of ما هي الاحتمالات per 100,000 words in Al-Masri Al-Yawm 2010.	CorpusName =Masri2010 word: mA hy AIEHtmAlAt search string: — ما هي الإحتمالات — mA hy AIEHtmAlAt database: Al-Masri Al-Yawm 2010 search time: 2 seconds part of speech: string search part of speech: string total number of occurrences: 0  0 instances of ما هي الإحتمالات per 100,000 words in Al-Masri Al-Yawm 2010.

Last, there is his use of verbal sentence structure. While English sentences start with a subject, an Arabic verbal sentence begins with the verb. In this article, verbal sentences begin with the subject. Examples are ( التجربة يمكن استنساخها في محافظة أخرى ، دينج ) ، شياو بينج تولى الحكم وعمره ٨١ عاما، مهاتير محمد أعادوه لرئاسة وزراء ماليزيا وعمره ٩١ عاما، الوظيفة تتطلب الاستقلال التام، الوظيفة قد يراها البعض طلبا مغامرا (أو صادما).

Another point about this author is that he is most likely a businessman or at least has his hands in the industry somehow. The argument for this proposition is that the use of الثروات التعدينية [mining/mined resources] is also rare even on the same newspaper. It is much more common to use الثروات المعدنية" [mineral resources]



The author is also seen as a fearless have-nothing-to-lose person. Although the idea he is presenting is critical and can put his career and loyalty at stake; yet he boldly uses first person



**VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A  
STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP**

Marwa Hassan Mohammed AlNagheya

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

singular voice to suggest it (أتصدر لها، أقترحها). He includes others as he switches to first person plural voice (i.e. inclusive we) in two cases: first when talking them into enforcing his idea (e.g. هنا لن ; نضيع وقتاً في اختراع نظم وقوانين جديدة. سنستعير النظم والقوانين المطبقة أما إذا فشلت التجربة. e.g. (فليس هناك الكثير لكي نبكى عليه). So, it is as if that we are all in this together, we can do it quick, but remember I am the one who started it all.

The author also combines a number of typing choices that mark his writing style. Of all enclosing marks, he only uses guillemets, pointed double angle quotation marks, to enclose information (e.g. «جمال حمدان» and «حاكم سيناء»). It takes eight keyboard strokes to produce a pair of these quote marks (alt + 1,7,4 for « and then another alt + 1,7,5 for ») when could have just easily used inverted commas or round brackets that require half as many strokes. Another salient marker is the way he uses full stops. Apart from guillemets and one question mark, no mark is used but the full stop. It is not used; it is actually overused. It is used 43 times. It does not just mark the end of a sentence, but it also splits clauses within a sentence (e.g. (أما إذا فشلت التجربة. فليس هناك الكثير لكي نبكى عليه), to separate items in a series (e.g. الوظيفة تتطلب الاستقلال التام عن (بيروقراطية القوانين السائدة. في الاستثمار. في استخدام الأراضي مثلاً). He makes use of Arabic diacritics as he types: more than 90 % the initial glottal stop أ is typed with Hamza above ء. This variety of the letter Alif is produced by holding the SHIFT keyboard key and simultaneously pressing H. Also, initial glottal stops with Kasrah (underbar below) ا are typed with the diacritic Hamza below the letter Alif. This variety of the letter Alif is made by jointly holding the SHIFT keyboard key and the Y buttons together with the keyboard set on Arabic. The flaccid Alif with the long vowel, madda, is also observed (e.g. (الأسبوية). These diacritics are not always observed by all writers; they are at times all typed just as a diacriticless ا).

The author also differentiates between Haa' and Taa' with the two dots over the knotted Taa'. He has the habit of using two

different keyboard keystrokes to set them apart. A distinguishing that I think discriminates his style is that he almost always places the Tanwīn nunation diacritics above the penultimate letter not the final post nasalized one (e.g. تَشْرَبًا، نَجَاحًا، بَاهِرًا، خَطًّا، مَلْعَبًا، مَفْتُوحًا). While the last letter on the Arabic alphabet is written in two different forms (i.e. ي and ى). Our questioned author always goes for the dotted variety. His hand is accustomed on hitting the D keyboard key. Two words also give notice to the keyboard key choices of that author (سئ and شيء). Every time he is using the word شيء, he makes use of three keyboard keys to form it (A, N, and X) which is grammatically true, though the ungrammatical variety is used in Al-Masry Al-Youm itself (see Figure 10).

## Conclusion

Given the high lexical density of the disputed op-ed, the text is seen as rich in information, an expected criterion of an opinion article. What is worth noting here is that the degree of richness falls in the range of informal writing that is easier to be understood as found by Fakhfakh (2015). The lexical, syntactic and stylometric features detected in the article allows for the hypothesis of the study at hand that every author has traceable features that can be counted as signature of their own even if they are behind a keyboard and veiled by an editor.

Comparing the profile arrived at after manual analysis of the disputed article with the real author behind the article (Diab declared himself to be the writer of the Newton article on April 12, 2020) allows for the following matching. He is an ageing (born in 1944) male (Salah Diab), well educated (raised in the English School of Cairo and graduate of engineering school) and of high class (he owns the newspaper and his grandfather used to own another Al-Jiha), a businessman (co-founder and CEO of the first Egyptian oil services company), observant of and inclined to American culture (owner of an American oil company, headed by Vice President Dick Cheney and had been accused of receiving American support for the founding of Al Masry Al Youm [Media Ownership Monitor Egypt]).

**VEILED AUTHORSHIP: CAN PEN-NAMED OP-EDS HAVE A  
STYLOMETRIC SIGNATURE THAT SURVIVES EDITORSHIP**

Marwa Hassan Mohammed AlNagheya

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

**References**

- Agbeyangi, A. O., Abegunde, O., & Eludiora, S. I. (2020, March). Authorship verification of yoruba blog posts using character n-grams. In 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS) (pp. 1-6). IEEE.
- AlHekaya (2020, April 17). رئيس تحرير المصري اليوم يرد على الهجوم على مقال (نيوتن) بخصوص سيناء [Video file]. Retrieved from [https://www.youtube.com/watch?v=i3W\\_FrlprT0](https://www.youtube.com/watch?v=i3W_FrlprT0)
- Anthony, L. (2023). AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>
- Argamon, S., & Levitan, S. (2005, June). Measuring the usefulness of function words for authorship attribution. In *Proceedings of the 2005 ACH/ALLC Conference* (pp. 4-7).
- Baayen, H., Van Halteren, H., Neijt, A. & Tweedie, F. (2002). An experiment in authorship attribution. JADT. 6es Journ'ees internationales d'Analyse statistique des Donn'ees Textue
- Coulthard, M., Johnson, A., & Wright, D. (2017). *An introduction to forensic linguistics: Language in evidence* (2<sup>nd</sup> ed.). Routledge.
- Eder, M. (2011). Style-markers in authorship attribution: a cross-language study of the authorial fingerprint.
- Fakhfakh, M. (2015). The readability of international illustration of auditor's report: An advanced reflection on the compromise between normative principles and linguistic requirements. *Journal of Economics, Finance and Administrative Science*, 20(38), 21-29.
- Grieve, J. (2007). Quantitative authorship attribution: An evaluation of techniques. *Literary and linguistic computing*, 22(3), 251-270.
- Grieve, J., Clarke, I., Chiang, E., Gideon, H., Heini, A., Nini, A., & Waibel, E. (2019). Attributing the Bixby Letter using n-gram tracing. *Digital Scholarship in the Humanities*, 34(3), 493-512.
- Halvani, O., Winter, C., & Pflug, A. (2016). Authorship verification for different languages, genres and topics. *Digital Investigation*, 16, S33-S43.
- Hunt, K. W. (1965). Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3.
- Hyryn, O. V. (2025). Potential of Authorship Verification Through Automatic Syntactic Analysis. *Zhytomyr Ivan Franko State*

- University Journal. Philological Sciences. Vol.1* (104), 126-133. DOI 10.35433/philology.1 (104).2025.126-133
- Jagaiah, T., Olinghouse, N. G., & Kearns, D. M. (2020). Syntactic complexity measures: Variation by genre, grade-level, students' writing abilities, and writing quality. *Reading and Writing*, 33, 2577-2638.
- Kestemont, M. (2014, April). Function words in authorship attribution. From black magic to theory?. In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)* (pp. 59-66).
- Michaelson, S., Morton, A. Q., & Hamilton-Smith, N. (1979). Fingerprinting the mind. *Endeavour*, 3(4), 171-175.
- Mining. (n.d.). In Oxford. com dictionary. Retrieved from <https://www.oxfordlearnersdictionaries.com/definition/english/mining?q=mining>
- Miri, K. (2020, April 14). Suspicious Deals. *Akhbar Al-Youm*. Retrieved from <https://akhbarelyom.com/news/newdetails/3032590/1/%D8%B5%D9%81%D9%82%D8%A7%D8%AA-%D9%85%D8%B4%D8%A8%D9%88%D9%87%D8%A9>
- Mosteller, F., & Wallace, D. L. (1963). Inference in an authorship problem: A comparative study of discrimination methods applied to the authorship of the disputed Federalist Papers. *Journal of the American Statistical Association*, 58(302), 275-309.
- Nini, A. (2018). Developing forensic authorship profiling. *Language and Law / Linguagem e Direito*, 5(2), 38-58.
- Nini, A. (2023). *A theory of linguistic individuality for authorship analysis*. Cambridge University Press.
- Nini, A., Halvani, O., Graner, L., Gherardi, V., & Ishihara, S. (2024). Authorship Verification based on the Likelihood Ratio of Grammar Models. *arXiv preprint arXiv:2403.08462*.
- Stamatatos, E. (2016). Authorship Verification: A Review of Recent Advances. *Res. Comput. Sci.*, 123(1), 9-25.
- Koppel, M. & Jonathan, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology* 60(1)9–26.
- Zhang, Y., Bumber, D., Hosseinia, M., Yang, F., & Mukherjee, A. (2021). Improving Authorship Verification using Linguistic Divergence. *arXiv preprint arXiv:2103.07052*.