

Was the president framed?

Language DNA, plagiarism and the incrimination of a president

Brief note on some linguistic aspects of the case of the *Prosecutor v Uhuru Muigai Kenyatta*
ICC-01/09-02/11

by Dr John Olsson

Forensic Linguistics Institute

www.thetext.co.uk

Abstract

This paper considers the witness statements used to incriminate President Kenyatta of Kenya in the recent International Criminal Court (ICC) trial. All of the witness statements analysed, supposedly written by people present at meetings or scenes of violence, were either fabricated or plagiarised, or both.

This is a serious indictment of procedures used in international criminal trials. The author, a forensic linguist and law graduate, has worked on previous international cases where similar problems were encountered.

It is not intended as an indictment of the ICC or other courts: prosecutors are placed under huge pressures to get results, and investigations become unwieldy very rapidly. However, if international justice is to work at all, the methods of collecting and assessing evidence need to undergo a radical, perhaps even cosmic, change for the better. If this is not done, international justice will simply become an academic topic and not a reality.

The author

Dr Olsson works as an independent forensic linguistics consultant and has written reports or given evidence to courts in approximately 500 civil and criminal cases in a number of countries, especially the UK. He has written a number of papers and textbooks on the subject. Dr Olsson is frequently consulted on civil (intellectual property, plagiarism, defamation, cyber-bullying, internet trolling, and authorship) and criminal (authorship of emails, cell

phone texts and social media postings, as well as anonymous hate mail and police statement fabrication) cases. He is also a law graduate.

Introduction

The Republic of Kenya, an important and prosperous east African country (and an early signatory to the Rome Statute, which established the International Criminal Court) was riven with ethnic violence in the aftermath of the 2007 presidential elections. In the course of time Uhuru Kenyatta, when he was a minister in the government, was accused by the ICC Prosecutor of being involved in the violence and despite this allegation he was later elected President. Charges were also brought against other persons, including William Ruto, the current vice-president and others. In this document, President Kenyatta will be referred to as 'the defendant'.

Background

In the matter of *Prosecutor of the ICC v Uhuru Muigai Kenyatta*, the Prosecutor, Ms Fatou Bensouda, withdrew all charges against the defendant, the President of the Republic of Kenya, on 5 December 2014. Nearly a year before, on the 19 December 2013 the Prosecutor had conceded publicly that: "the case against Mr Kenyatta does not satisfy the high evidentiary standards required at trial". This statement arose as the last remaining key factual witness had admitted in a Prosecution interview that he had lied about the involvement of the Defendant in the election violence. Later, Ms Bensouda averred that the government of Kenya was not co-operative in its dealings with the court and had, effectively, obstructed the prosecution by hampering access to important government documents which, it is said, would have enabled the court to form an opinion as to the daily movements of the Minister (as Mr Kenyatta, now the President, then was) in regard to specific locations linked to incidents related to the incitement of violence in the aftermath of the 2007 elections.

While no comment will be made by this author on those matters, there is another aspect to the case which was not mentioned by the prosecutor.

The difficulties of international criminal prosecution

Before outlining that aspect of the case, the author asks the reader to note that this paper is not in any way intended as a commentary on the correctness of the prosecution or otherwise. That is strictly a matter for the prosecutor, the court and the defendant's legal advisers. The events in Kenya in 2007 and 2008 represent an epic human catastrophe for its people, and it is right that where there is reason to believe that oppression, whether by violence, removal of populations, or through any other means of denying people fundamental rights, may have been orchestrated or condoned by a government, organisation or an individual, then inquiries should be made as to who is responsible. In matters of this nature, the prosecution has the most daunting and unenviable task. Any reasonable person, and any person with democratic values would, and should, salute the court for attempting to provide justice for the victims of the Kenyan tragedy. That this was the prosecutor's sole motivation in the matter is evident from her and her predecessor's many declarations in this regard and their undoubted passion for the cause.

However, previous international cases in which the author has been involved as an expert, have demonstrated the difficulties of obtaining reliable, independent testimony from witnesses of mass crises. Invariably, time and the consequences of trauma vitiate human memory, and factional interests begin to prevail: in some cases, an opportunity to settle old political scores emerges. For those reasons, any prosecutor faced with the overwhelming task of attempting to collect evidence deserves credit simply for making the effort. After a team of advisers and investigators, international and national, is dispatched to carry out the task, local sources of information have to be relied on. The task is doubly complicated by the fact that, even if dependable accounts can be obtained, witness statements have to be translated. In a Rwandan case some years ago, statements were translated from Kinyarwanda (in all probability standardised from a variety of local dialects and variants) into French and subsequently into English and a number of serious errors in the translation were found by the author and others. The task in Kenya in the present case was no different in linguistic complexity.

Witness statements

In the present matter, a large number of witness statements and other documents were submitted to the court in connection with the post-election catastrophe. Certain of these were

to be used as evidence against the defendant, Uhuru Kenyatta. It rapidly became clear to the author, however, that there were many linguistic similarities between the statements and that they were not the independent eye-witness accounts they purported to be.

Plagiarism

The working hypothesis used by the author was that if two statements were of common authorship they could bear the kinds of similarity to each other found in cases of, for example, academic plagiarism.

Plagiarism has been the object of study by a number of forensic linguists in recent years, and in this connection Professor Malcolm Coulthard and Dr Alison Johnson (Coulthard and Johnson 2008) made a prescient observation, namely that plagiarism – unlike certain other indicators of common authorship – did not require large quantities of comparable language to be established: “[W]e can assert that even a sequence as short as ten running words has a very high chance of being a unique occurrence” (Coulthard and Johnson 2007: 198). Coulthard and Johnson give the example of “I asked her if I could carry her bags”, a nine word string. When searching for the string “I asked her if I could” at the time of writing 7,700 instances were found on the internet. When the word ‘carry’ was added only seven instances were found. This shows that when a generic-sounding phrase ‘I asked her if I could’ is added to a lexical word with contextual potential such as ‘carry’, the frequency of the phrase’s usage diminishes exponentially, even hyper-exponentially. No internet instances were found when the search was for the full string, ‘I asked her if I could carry her bags’.

The author had reached a similar conclusion, at about the same time that Professor Coulthard published his seminal paper on the idiolect (Coulthard 2004). The two forensic linguists had arrived at the same conclusions, independently of each other, at more or less the same time (see Olsson 2004: 112-114).

The finding was, essentially, that even a very short phrase consisting of words known to most users of the language, could be unique to that person. The implications of this for the detection of authorship are stunning, and Coulthard and Johnson noted that results of this type might be comparable with the kinds of results produced in court by DNA experts (Coulthard and Johnson 2007: 198).

In the witness statements submitted to the ICC in the Kenyatta case, many instances of the same phenomenon were found, a few examples of which are given below.

Practicalities of statement-making in the international context

It is important to emphasise that in any investigation witness statements – whether written or audio or video-recorded – should be quarantined from each other. This does not just involve making sure that contact between witnesses is kept to a minimum (insofar as is practicable), but also that investigators and those conducting interviews need to avoid re-using phrases and expressions previous witnesses or suspects have used. As an investigation gets underway and more information is obtained about possible involvement by suspects, the temptation is to piggyback evidence. Investigators get into a certain way of asking questions and conducting interviews, and corners are cut. One result is that words and phrases from one witness are recycled by investigators and previous items of evidence become apparently confirmed. It is all too easy to infer similarities from two apparently parallel accounts: the problem is that the accounts are probably nowhere near as similar to each other as the investigators think. All that has happened is that the investigators themselves have polluted the inquiry by, perhaps unwittingly, feeding ideas back to successive groups of witnesses. The evidence is not generated *sua sponte* by live witnesses. Rather, independent accounts are occluded by unnecessary investigator-interventions, and the evidence becomes cloned.

The golden rule is that statements need to be in the words of the statement maker. The interviewer's language must be careful and neutral. This involves planning much more than just the information to be probed, but the words in which the questions should be asked. Where interpreting is involved, later checks by supervising translators need to be carried out. Both investigative and translation teams need to be rotated. In rural areas in Africa, it will not be uncommon for investigators to have to put together a team of translators competent not just in the *lingua franca* of the area, but they will also have to obtain the services of interpreters of several distinct languages, as well as a possible myriad of local dialects. Gender and age are sensitive subjects, because of the complex respect culture. In Bantu languages some words will be taboo in a cross-gender interaction. The difficulties of interpreting in these circumstances are far outside the experience of those accustomed to interpreting in the European context. Once documents are sent to be typed, they should be carefully checked, but any alterations or suggestions must be recorded. It is better to have one

good statement, properly obtained and recorded, than twenty or thirty poor ones. The statements in this case were low on meaningful content and high on innuendo, gossip and rumour.

Examples of some similarities across the documents

Because the documents in this case are still confidential, they have been heavily redacted for identity purposes and in no case can the writer of the document be identified. The reader will appreciate that at the time of writing, it is still necessary for the author to be discreet. For this reason, only a few examples of similarities will be given. All names have been removed, including the names of organisations.

Example 1

This pair concerns the claim that several people were sent to see a certain prominent person to persuade that person to take part in a certain activity.

A:

(ABC)¹ **told us that they had been sent and were representing X, Y, Z and A**².

B:

The XYZ³ side **told us that they had been sent and were representing X, Y, Z and A**⁴.

¹ redacted

² Names redacted

In the above pair, there is an identical common 12-word string in each example.⁵ Above, it was explained that a common string of this length is highly unlikely to arise by co-incidence. The probability of the two sentences, and hence the statements which contained them, being by two different authors acting independently of each other, is extremely low.

What makes the parallel even more compelling is the absence of the preposition ‘by’ (i.e. ‘had been sent by’) which is common to both sentences. It is accepted in academic plagiarism investigations, and in police statement fabrication cases, that an error common to two samples – such as an error of fact, or a grammatical, spelling, or idiomatic error – diminishes the possibility of co-incidence even further: even a basic error, will have potential significance. When the same error is found in an identical run of words greater than six or seven words (or even less) credibility as to the two phrases arising independently of each other can be suspended with reasonable certainty.

The author considered whether ‘they had been sent’ was in any sense a fixed phrase. Indeed, with many millions of results on the internet it is certainly a very common phrase, although the use of the simple past tense, ‘was/were’, is more common than the past perfect tense, ‘had been’. However, when preceded by ‘told us that’, and followed by ‘and’, giving ‘told us that they had been sent and’, there are only three internet results (using a well-known search engine). The omission of ‘by’ and the inclusion of ‘and’ is what causes this dramatic reduction in frequency. This is because variations of ‘to be sent’ are about forty times more likely to be followed by ‘by’ than by ‘and’. The writer knows the reader will be seeking information about the agent of the ‘sending’, and the syntactic link to agency will be through use of the preposition ‘by’.

Hence, in Example 1 above, that omission plays a crucial role in showing the common authorship of the two purportedly independent sentences (note: all phrasal searches on the internet are encapsulated by double quotation marks). Note also the identical sequence of the persons referred to (names redacted).

³ redacted

⁴ Names redacted

⁵ Only ten words of the 12-word sequence are shown here.

Example 2

In this pair, the supposedly different authors of two allegedly unconnected statements are claiming the participation of the defendant in a series of ethnic murders and other atrocities.

A:

I believed all the time that the revenge attacks had the blessings of X, Y, Z and A. ⁶

B:

All the time that.... on the understanding that the revenge attacks had been authorized and had the blessings of X, Y, Z, and A. ⁷

In this pair, the common string is ‘that the revenge attacks (...) had the blessings of X...’. No internet document was found containing both phrases “that the revenge attacks” and “had the blessings of (etc)”, either sequentially or as separate strings.

Note also the identical sequence of X, Y, Z and A. It is also interesting that the same names, in the same order as in Example 1 are given in this example. Two of these names are given with first names – the same two names. Thus, the supposedly independently produced statements in this example are connected with the supposedly independently produced statements in Example 1. Recall that each statement, not just each pair, must be independent of every other statement.

A curious detail of both examples here is that each contains a mental projection: ‘I believed all the time that’ or ‘on the understanding that’. That is to say, both reflect an opinion or a belief or a judgment, in a separate, projected clause. Crucially, these projections are both time-related using the same adverbial phrase, namely ‘all the time’.

⁶ Names redacted

⁷ Names redacted.

Example 3

The third and final example in this brief paper is a general commentary on the claim that a certain person must have known about the atrocities.

A:

A policy of extermination on this scale could not possibly be formulated without clearance from Z. In any event, Z cannot feign ignorance over the extra judicial executions. He reads newspapers, watches television, listens to FM radio stations and receives [something] from [somebody]

B:

Extra judicial executions and forced disappearances of youths in excess of 7,000 can only happen with a policy in place approved by Z himself. In any event, he reads papers, watches television, listens to radio and has [something] from [somebody] and so he has been aware of the extra judicial executions and their extent.

In this pair, the collocation of (news)papers, television and radio is very common and therefore not remarkable. What makes the above sequence interesting in terms of authorship is the addition of the ‘something’ (the same ‘something’) from ‘somebody’ (the same ‘somebody’, but phrased slightly differently).

Even that, however, would not necessarily be unusual until we consider several further correspondences, including the innocuous looking qualifier ‘in any event’ as well as the presence of ‘policy’, ‘Z’ (a person) and ‘extra judicial executions’. These serve to add an authorial dimension to the parallel.

Crucially, the seven word string “he reads papers, watches television, listens to” is not found on the internet (in this regard the difference between ‘papers’ and ‘newspapers’ is trivial and may be disregarded for the purposes of comparison – the phrase does not occur, even with ‘newspapers’ substituting for ‘papers’).

Conclusion

In this very brief paper the author has attempted to do no more than show some of the similarities between supposedly independently produced witness statements in a major international criminal case. Similarities of this nature and extent were found across more than thirty of the statements in the case.

Other statements contained allegations as grave as those in the above examples. However, a feature of the case not found in police statements or witness accounts in other cases, was the presence of speculation, rumour and innuendo, as noted in the examples above. In all of the world's criminal legal systems (different rules apply to civil cases in some jurisdictions) such features are rigorously excised from allowable testimony. The opinion or belief of the witness is rarely permitted (other than, necessarily, that of the expert witness). This is for the simple reason that what any criminal court wants to know from a witness is what s/he saw, and what s/he heard, and what s/he did – not what s/he believes, or heard *about*, or – as in many cases in the documents, '*heard rumours about*' (or phrases of that character).

However, even in the few examples given here, four instances of this phenomenon were found: "ABC told us", "the XYZ side told us", "I believed all the time", "all the time....on the understanding that". In one particular statement, a lengthy claim as to the presence of the defendant was made in the active indicative tense. It gave the clear impression of being an eye-witness account. It was only towards the end of a description which ran to several paragraphs that the witness gave the qualification that he had heard or been told of the events he was purporting to describe and that he had not, in fact, seen the defendant at all. Needless to say, the person claimed to have seen the defendant at a particular place at a particular time, connected with the atrocities.

It is also important to note that where we see minor differences between examples – for instance, 'newspapers' instead of 'papers', and 'FM radio stations' instead of 'radio' in Example 3 above, these will often indicate an effort on the part of the plagiarist to conceal the similarity. The same technique is used by persons fabricating police statements, as well as by those in academia who plagiarise.

As a result of the above analysis, the author concluded that a concerted effort had been made to implicate the defendant. Certain characteristics of the documents indicated that they were linked to each other linguistically. In Examples 1 and 2 above, the sequence of names and the forms in which they are presented were nearly identical. There were many other types of

semantic and lexical linking across the documents, showing the involvement of one particular individual, who was able to be profiled with considerable accuracy.

The burning question is why this was allowed to happen. Why were these matters not picked up at an earlier stage?

Quality of evidence

It is absolutely right that a suspect against whom credible evidence is available, should be challenged in a criminal court, whoever he or she is. However, as is patent from even just the above examples, the evidence against the defendant in this case was manufactured. Even a cursory glance at such evidence by a competent forensic linguist would have revealed at the very least a cause for suspicion. We may ask ourselves why proper scrutiny did not take place on this occasion.

The world is not immune to civilian population crises of the type experienced by hundreds of thousands of Kenyans in 2007. It is entirely understandable that the international community seeks to uncover the perpetrators. Too often, however, the clamour of the world places undue pressure on those involved in the investigation. It is easy to yell from the sidelines, but not so easy for those 'in the thick of it'.

As indicated earlier, the difficulties of gathering accurate and reliable evidence of culpability cannot be underestimated. There is a multiplicity of causes: it may be difficult to locate traumatised victims; infrastructures such as roads and telecommunications will have been disrupted, or even destroyed; new power dynamics may have come into force, and, quite naturally, victims will be seeking vengeance. More widely, opportunistic individuals may seek to exploit the misery of their compatriots – as happened in this case. The opportunities for fabricating evidence multiply – in proportion to the lapse of time between the event and its investigation; the number of participants involved; the various languages and dialects from which translations are required to be made; the civil disorder to the region, and the numbers of persons displaced, killed or assaulted, as well as many other factors. From a superhuman effort to gather evidence, a number of statements will emerge, often obtained under extraordinarily difficult circumstances. The point is, no matter how hard it may have been to collect them, unless each statement is independently produced, faithfully, in the words of the statement maker, without influence or contamination by any person, it cannot be depended

upon. This is a lesson not entirely learned even now in the so-called advanced democracies, although there are fewer perversions of justice in this regard than there used to be.

It is also important to ensure that, as far as evidence gathering is concerned, persons with a vested interest in the outcome must be kept out of the equation. In any international inquiry certain local faces will keep ‘popping up’ to ‘assist’ the international investigation team, when in reality their purpose may be less than altruistic.

Linguistics provides a straightforward and reliable way of testing for plagiarism, fabrication, collusion and incrimination. Bodies of the stature of the International Criminal Court, as well as national and international investigation agencies, should consider seeking the assistance of forensic linguistics, a well-respected academic subject, now taught at a number of UK universities, and online at the long-established Forensic Linguistics Institute. Its usefulness has proved itself in a number of courts throughout the UK, the USA and other countries.

The world is not becoming a safer place. Disasters, both of the natural variety and those contrived by humans, will continue to happen. It is important that the authorities are able to weed out unreliable evidence at an early stage. Careful gathering of evidence may mean it takes longer to bring people to justice, but the international community must learn to accept that simple fact. The alternative is no justice at all.

References

Coulthard, M. 2004. *Author Identification, Idiolect and Linguistic Uniqueness*. *Applied Linguistics* 25, 4, 2004, 431-447.

Coulthard, M. and Johnson, A. 2007. *An Introduction to Forensic Linguistics: Language in Evidence*. London: Routledge.

Olsson, J. 2004. *Forensic Linguistics: An introduction to Language, Crime and the Law*. London: Continuum.