A European language detection software to determine asylum seekers' country of origin:
Questioning the assumptions and implications of the EUAA's project

Cecilia Manzotti

PhD candidate, School of Law
University of Sussex (UK)

Abstract

Since 2017, Germany has used an Artificial Intelligence (AI) language recognition system to verify asylum seekers' claims regarding their country of origin before the asylum interview. Inspired by the German experience, in 2023 the European Union Asylum Agency (EUAA) announced a project to develop a European language detection software for determining asylum seekers' country of origin. The tool will be made available to all EU member states and complemented by a second-line pool of language analysts. This blog post raises attention on the EUAA's project and questions the assumptions on which automatic language indication is based as well as its implications on the overall credibility assessment of asylum seekers' nationality claims and, more fundamentally, on the right to seek asylum.

1. The EUAA's project to establish a European language detection software for determining asylum seekers' country of origin

In 2022, the [European Union Asylum Agency (EUAA)](#) reported that seven European Union (EU) member states (Austria, Germany, the Netherlands, Romania, Denmark, Sweden and Finland) and Switzerland, have used Language Analysis for the Determination of Origin (LADO) as a tool for establishing asylum seekers' country of origin for several years, and six other member states (Croatia, Greece, Malta, Polonia, Portugal and Slovakia) are considering introducing LADO in the near future. Based on the assumption that the way a person speaks reflects their *origin* – a notion which is often conflated with *nationality* – LADO consists in the analysis of speech samples to ascertain whether the speaker really originates from the country they claim to be from. More rarely, LADO is used to verify whether asylum applicants genuinely originate from a specific region within a country or belong to a certain ethnic group. LADO and its use in asylum procedures have fuelled lively debates, especially among linguistics. While LADO is generally conducted by humans, since 2017 Germany – until now the only country in Europe – has implemented an Artificial Intelligence (AI)[1] tool to detect the languages or dialects spoken by asylum applicants.

Inspired by the German experience, the EUAA has recently launched a project to establish a «common European platform to identify the country of origin of applicants through language assessment», including a first-line AI language detection tool and a second-line pool of language analysts. The project was announced in the [2023 EUAA's Strategy on Digital Innovation in Asylum Procedures and Reception Systems](#), which to date is the only publicly available official document providing information on the initiative. Under the project, whose implementation timeline spans over a period of 10 years, the EUAA will initially foster coordination between stakeholders and develop common standards and procedures. At a later stage, once the EUAA's mandate is adjusted as needed, the agency will establish an AI language detection tool, create a pool of analysts, and build the capacity of a team in charge of the system. In the EUAA's view, the effort will reduce costs related to LADO, level differences among national administrations in terms of digitalisation, «facilitate more efficient and smarter identification» of asylum seekers, «ultimately leading to better and faster decisions» (EUAA's *Strategy on Digital Innovation*, pp. 28-30).

Although the EUAA's project may have a significant impact on the substance of asylum decisions and procedural guarantees for asylum applicants, so far it has gone virtually

---

[1] For the purpose of this blog post, I adopt the definition of 'AI system' contained in Article 3(1) of the [EU AI Act](#). Under certain circumstances, the use of AI for identifying asylum seekers who are unable to prove their identity is permitted under the AI Act (recital 33).

unnoticed both among experts and in the media. Therefore, this contribution draws attention on the EUAA's project and investigates its potential impact on the credibility assessment of asylum seekers' claims regarding their country of origin under the Pact on Migration and Asylum (the Pact). Since there are several reasons to believe that a European AI language detection tool will be largely based on the software used by Germany – which Germany has actively promoted across Europe and the EUAA described as «providing fast and reliable assessment» of asylum applicants' origin (EUAA's *Strategy on Digital Innovation*, p. 28) – this blog post will start by critically examining the German automatic language detection system. Based on the German experience, the blog post will then attempt to foresee how the EUAA's project will be operationalised in the context of the new or amended procedures introduced by the Pact. This contribution questions the assumptions on which automatic language analysis is based as well as its implications on the overall credibility assessment of asylum seekers' nationality claims and, ultimately, on the applicants' right to seek asylum.

2. *The German automatic language and dialect detection system*

Following the 2015-2016 influx of asylum seekers and under the pressure of the scandal of Franco A., a German far-right extremist who was granted subsidiary protection pretending to be Syrian, in 2017 the German Federal Office for Migration and Refugees (*Bundesamt für Migration und Flüchtlinge*, BAMF) introduced three AI-based identification systems. These include an automatic language and dialect recognition system (DIAS, *Dialektidentifizierungsassistent*), a software for the transliteration of Arabic names that also indicates how often a name's spelling is used in the applicant's claimed country of origin and in other countries, and a software that analyses data stored in asylum seekers' electronic data carriers for the purpose of establishing their identity and/or nationality. Since 2018, the three identification systems have been used as a standard procedure during the registration of asylum applicants in all BAMF branch offices and reception centres, if the person cannot provide a valid passport or passport substitute[2] or if the document's authenticity cannot be immediately confirmed.

Concerning DIAS, in particular, following the registration of personal data, the BAMF's officer dials an internal phone number and the applicant is asked to describe a picture or discuss freely a topic in their native language on the phone for around two minutes. Their speech is recorded and analysed by DIAS, which automatically produces a report indicating the language(s) or dialect(s) spoken by the applicant and with which probability. The report is then used by the BAMF officer to prepare specific questions about the applicant's origin for

---

[2] The BAMF's *Instructions on Identity Verification in Asylum Procedures* (*Dienstanweisung Asylverfahren, Identitätsfestellung, 2023*) define a 'passport substitute' as «a document that, alone or with a visa or residence permit, authorizes cross-border travel and fulfils some, but not all, of the functions of a passport. In particular, the identity card (ID card) is relevant in this context». (Translation from German into English revised by the Editorial Team).

the asylum interview. If the report's conclusions contradict the applicant's claim regarding their country of origin, the applicant must be informed and given the opportunity to respond. As of November 2023, DIAS is used for the five major Arabic dialects (Egyptian, Gulf, Iraqi, Levantine and Maghrebi), Dari and Farsi. A Pashto language model was introduced in 2022, although it is unclear whether it is still in use, and a language model for Kurdish has been in the pipeline for a couple of years.

The BAMF justifies the use of DIAS based on Section 15 of the German Asylum Act that sets out the applicant's general obligations to cooperate, and Section 16(1), which specifically provides that oral statements can be used to determine a person's country or region of origin provided that the person was informed beforehand. The Federal Government pointed out that «[t]he results of the dialect recognition can neither confirm nor refute the information on origin» but only «provide an indication of the applicant's origin, which is taken into account in the context of the interview, which also serves to clarify identity and nationality».[3] In very exceptional cases, if doubts concerning the applicant's country of origin persist after the asylum interview, a «Speech and Text Analysis», carried out by external linguists based on a new and longer speech sample, may be recommended.[4]

According to the Federal Government, the use of DIAS has several benefits, including the verification of asylum seekers' origin early on in the procedure, the fact that additional data are made available to decision-makers to support asylum decisions, an overall acceleration of the procedure, a reduction of fraud and increased security. On the other hand, linguists, civil society actors and members of the Parliament have expressed serious concerns regarding DIAS' accuracy and reliability. Indeed, the Federal Government reported a language recognition rate of 80% for Arabic dialects in 2017, which increased to 87% in 2023, and 75% for the other languages in 2022. This means that the software provides a wrong result for around 20% of the applicants who undergo the procedure. Other criticisms concern the BAMF's lack of transparency regarding the software's algorithms and the language samples distribution, the risk of self-perpetrating bias, and the government's failed promise to commission an independent evaluation of the system. Leaving in the background these issues, which have already been explored in some of the literature on the use of AI in asylum procedures (see, for example, D. OZKUL, *Automating Immigration and Asylum: The Uses of New Technologies in Migration and Asylum Governance in Europe*, 2023), the use of a language detection software to ascertain asylum seekers' country of origin poses fundamental questions regarding the evidence and standard of proof for assessing the credibility of asylum applicants' nationality claims.

---

[3] Translation from German into English revised by the Editorial Team.
[4] On the difference between 'language indication' and 'language analysis', see EUAA, *Executive Summary, Study on Language Assessment for Determination of Origin*, September 2022, p. 13.

First, the BAMF's guidance problematically describes DIAS as a tool to establish an applicant's *nationality* or *country of origin*, which is defined in EU refugee law as a person's country of nationality or, in the case of stateless persons, country of former habitual residence (recast Qualification Directive, article 2(n); Qualification Regulation, article 3(13)). Although the guidance available in other countries using LADO or developed by the EUAA (p. 36) specifies that «[l]anguage analysis does not reveal the country of nationality of the applicant as such, but the place (or one of the places) where the applicant has socialised by residing there for a longer time and interacting with the community», in asylum decisions LADO conclusions are generally used – either openly or implicitly – to determine a person's nationality. Indeed, what is relevant when it comes to the assessment of asylum seekers' risk of persecution or serious harm under refugee law is the concept of *country of nationality*, and not the notion of *country of socialisation*. But what does constitute evidence of a person's nationality in the absence of any identity or travel document? While asylum authorities generally resort to LADO and questions to test the applicant's knowledge of their alleged country of origin, neither language nor knowledge of a country's geography and traditions constitute evidence of nationality. Nationality is a legal status, which may or may not correspond to a person's main country of socialization. Accordingly, more pertinent questions would concern, for example, identity and travel documents issued by the applicant's alleged state of nationality, the applicant's attempts to obtain these documents, the applicant's place of birth and their parents' nationality, and the applicant's access to the rights and entitlements reserved to nationals of the state in question.

Second, and related to the previous point, the fact that DIAS is used when a person's asylum application is registered, that is before the asylum interview, means that language indication is given priority over the applicant's testimony as evidence of nationality. Indeed, the results produced by DIAS orient the asylum interview, and not the other way around, and it is hard to believe that they do not create a prejudice in the interviewer. The use of DIAS before giving the applicant the opportunity to explain in detail their personal situation and reasons for seeking asylum, even in the absence of any negative credibility indicators, appears even more questionable considering the software's poor reliability. Additionally, although the BAMF has repeatedly specified that DIAS only provides an indication of the applicant's nationality, it remains unclear what evidentiary weight language indication should be given in the overall credibility assessment of the applicant's nationality claim. In the absence of any clear guidance, the question risks being left to the discretion of individual decision-makers.

Third, a closer look at the conditions that trigger the use of DIAS shows that asylum applicants are expected to substantiate their nationality to a standard of proof higher than the *balance of probabilities* generally required in refugee status determination. Indeed, the BAMF's instructions provide that DIAS must be used if the applicant's identity and nationality cannot be *proved*, *determined with certainty* or *established beyond doubts*. On the other hand, the BAMF seems to adopt a much more generous standard of proof when it comes to its own

determination of asylum seekers' nationality through language indication, considering DIAS' low accuracy level. The fact that DIAS is used whenever an applicant does not produce any valid passport or passport substitute or the BAMF doubts the documents' authenticity also constitutes a weakening of the principle – included in German law (Residence Act, Section 5(3)) and well-established in EU refugee law (Qualification Directive, article 4(5); Qualification Regulation, article 4(5); ECtHR, *F.N. and others v. Sweden*, paragraph 72) – that asylum seekers do not need to produce a passport to substantiate their identity and nationality.

In sum, the German automatic language recognition system relies on problematic assumptions regarding the relationship between language and nationality. Furthermore, its systematic use before the asylum interview whenever an applicant fails to provide a valid identity or travel document or the BAMF doubts the authenticity of the applicant's document reflects the adoption of an unduly high standard of proof for the applicant. Moreover, it results in automatic language detection being given priority over the applicant's testimony in the credibility assessment of nationality claims, despite DIAS' inaccuracy and controversial nature.

3. *Where does the EUAA's project fit within the Pact?*

Under the Pact, the determination of asylum seekers' country of origin is not only critical to the assessment of their fear of persecution or serious harm but may also substantially affect the level of procedural guarantees to which applicants are entitled (C. MANZOTTI, *Nationality Status Determination in Asylum Procedures under the CEAS and the Potential Impact of the 'New Pact on Migration and Asylum', 2023*). Indeed, following a pre-entry screening, applicants who originate from countries with a low recognition rate and those who do not cooperate in the identification procedures, notably concealing their nationality, will be channelled into the asylum border procedure. The latter implies fewer procedural safeguards compared to the regular examination procedure and practical restrictions that may substantially affect the applicants' ability to assert their claim. Under the new rules, the nationality of applicants for international protection is determined and recorded for the first time during the pre-entry screening, which includes preliminary health and vulnerability checks, identification, registration of biometric data and security checks (Screening Regulation, article 8(5)). The screening authorities must include an «indication of nationalities or statelessness» (article 17(1)(b)) in the screening form, specifying if the information recorded has been «declared by the person» or «confirmed by the authorities» (article17(3)).

Considering that the purpose of the pre-entry screening is to ensure that asylum applicants «are referred to the appropriate procedures at the earliest stage possible and that those procedures are continued without interruption or delay» (recital 7), it is reasonable to foresee that the automatic language detection tool that the EUAA is going to develop will be made

available to the screening authorities. Although the Screening Regulation does not mention language indication or analysis anywhere, Article 14 includes «data or information provided by or obtained from the applicant» among the types of evidence that should be used to establish or verify asylum seekers' nationality, together with identity, travel or other documents and biometric data. The phrasing «data or information provided by or obtained from the applicant» is sufficiently broad to include the recording and analysis of a speech sample through a language recognition software. On the contrary, the EUAA's pool of language analysts is likely to come into play during the asylum procedure, since the pre-entry screening must be finalised within seven days at the border and three days within the territory of member states (article 8). As in the case of Germany, it is foreseeable that asylum authorities will be able to request an in-depth language analysis if doubts regarding the applicant's country of origin persist.

Since the goal of the pre-entry screening is to identify applications that are likely to be inadmissible or unfounded and channel them into the accelerated and border procedures as soon as possible, the results produced by the language detection software can be anticipated to be decisive in establishing applicants' country of origin in the absence of any valid identity or travel document. Importantly, the Screening Regulation does not foresee any possibility for the applicant to challenge the authorities' recording of their personal data during the screening. Although Article 17 provides that the applicant shall have the possibility to indicate that the information included in the screening form is incorrect and the authorities must record this, under the Regulation the applicant's view does not have any impact on the screening process and, given the short time of the procedure, is unlikely to lead to a more in-depth assessment at this stage. This means that the decision to examine an asylum application through a sub-standard procedure, severely limiting the applicant's rights, would largely be based on the results produced by an AI language recognition software that is grounded on wrong assumptions regarding the relationship between language and nationality and has proved to be inaccurate. The results of the automatic language recognition may also potentially affect the assessment of the asylum application, since [evidence](#) shows that amending a nationality record during the asylum procedure can prove extremely complex. This would be even more complicated for asylum seekers who have been identified as originating from countries with a low recognition rate or safe countries of origin and whose application has been rejected as unfounded or manifestly unfounded, because their appeal against the first instance decision would not have automatic suspensive effect ([Asylum Procedure Regulation](#), article 68(3)).

Ultimately, the use of an AI language recognition tool to establish asylum seekers' country of origin during the pre-entry screening would not be without consequences on the credibility assessment of asylum seekers' nationality claims, as the case of Germany has shown. Moreover, combined with the systematic channelling of applicants from certain countries of

origin into the border procedure, the use of automatic language indication would compromise the applicants' right to seek asylum. With their asylum applications being examined at the border and under limited procedural guarantees, applicants identified as originating from certain countries would face significant obstacles asserting their claim and challenging the first instance authorities' decision.

FURTHER READINGS

EMN-OECD, *The Use of Digitalisation and Artificial Intelligence in Migration Management*, February 2022.

D. EADES, *Nationality Claims. Language Analysis and Asylum Cases* in M. COULTHARD AND A. JOHNSON (eds), *The Routledge Handbook of Forensic Linguistics*, 2010.

H. HAHN, *Digital Identification Systems and the Right to Privacy in the Asylum Context: An Analysis of Implementations in Germany*, 2021.

K. WILSON AND P. FOULKES, *Borders, Variation, and Identity: Language Analysis for the Determination of Origin (LADO)* in D. WATT AND C. LLAMAS (eds), *Language, borders and identity*, Edinburgh University Press, 2014.

To quote this contribution: C. MANZOTTI, *A European language detection software to determine asylum seekers' country of origin: Questioning the assumptions and implications of the EUAA's project*, ADiM Blog, Analyses & Opinions, November 2024.