The International Journal of Speech, Language and the Law

IJSLL (PRINT) ISSN 1748-8885
IJSLL (ONLINE) ISSN 1748-8893

Article

Duly verified? Language analysis in UK asylum applications of Syrian refugees

Yaron Matras

Abstract

This article examines current LADO (Language Analysis for the Determination of Origin) practice as documented in a sample of recent reports that assess the speech of Syrian asylum seekers. I show how a method that purports to ensure transparency of data, and contextualisation of the analysis in published research, presents certain problems by framing the question of linguistic background as two juxtaposed hypotheses. I suggest that this approach risks biasing the results by excluding the possibility that speech can be varied and multi-layered. At the same time, points of concern around LADO procedures raised by researchers, such as analysts' inadequate qualifications and failure to take variation into account, are resonating with the judicial system via arguments presented in counter-expertise reports. I argue in favour of an inductive approach to language analysis that draws on a holistic dialect-geographical assessment while giving consideration to sociolinguistic and discourse analytical dimensions.

KEYWORDS: LADO, ARABIC, KURDISH, ASYLUM, MOBILITY, VARIATION, LINGUISTIC REPERTOIRES, DIALECT GEOGRAPHY

Affiliation
University of Manchester, UK
email: yaron.matras@manchester.ac.uk

IJSLL VOL 25.1 2018 XXX–XXX https://doi.org/ 10.1558/ijsll.35710
©2018, EQUINOX PUBLISHING

1. Background

Language Analysis for the Determination of Origin (LADO) has been applied by governments since the early 1990s to assist in the assessment of the legitimacy of applications for asylum (cf. Reath 2004; Eades 2005; Patrick 2012). The procedure rests on the assumption that individuals' speech is indicative of their place of origin. LADO is a relevant tool especially in cases where an applicant's right to asylum is undisputed if they are able to prove that they come from a country that is listed at the relevant point in time as unsafe. Syria has been one of those countries since the outbreak of the civil war in 2011. In this article I assess the reliability of a corpus of LADO reports carried out on the speech of Syrian refugees in the UK between 2015 and 2017.

In the period between 2011 and 2015, the total number of applications in the UK by refugees claiming to be from Syria was reported to be 5,465, of which 1,114 were refused by the UK Home Office. Of those, 173 applications were allowed following an appeal to the courts, while 238 appeals were dismissed.[1] For 2016, the UK government reported a total of 1,409 asylum applications from Syrian nationals. In the year ending in March 2017 the grant rate for both asylum applications and alternative protection measures for Syrian nationals was 86%, higher than for any other group of nationals.[2] Those rejected included cases where applications were referred to another EU state in which the applicant was thought to have stayed prior to arrival in the UK, as well as cases 'where it was found that the applicant did not hold Syrian nationality'.[3]

It is not known exactly what proportion of applications undergo a language analysis – either in general or specifically in cases of refugees claiming to be from Syria. In February 2013 the UK government provided authorisation to the UK Border Agency to use linguistic analysis in any case of doubt of the nationality of applicants claiming to be Syrian, Kuwaiti or Palestinian.[4] It is possible that the number of language analyses increased following that directive. Until 2014, the UK Home Office had contracted the company Språkab based in Sweden to carry out LADO reports. Coinciding with criticism of Språkab's practices (see Craig 2012; Campbell 2013), including a Supreme Court ruling from May 2014[5] (see Patrick 2016), the Home Office began to use an alternative supplier, Verified, also based in Sweden (Home Office 2017: 5). Between 2014 and 2017, altogether 4,760 language analysis reports were commissioned by the UK Home Office to Verified and Språkab, at a total cost of c. £1.2 million.[6] While the method of eliciting language data for analysis appears to have remained the same, the approach to data evaluation and the presentation of the findings in Verified reports differ significantly from Språkab's earlier practice.

My aim in this article is to review the method used by Verified in recent analyses of the Arabic and Kurdish speech of Syrian asylum applicants in the UK.

Drawing on a corpus of 50 recent reports by Verified obtained from casework material through legal counsel and anonymised, I discuss methodological problems pertaining to the use of published sources, the historical interpretation of features, variation and style shift, and the weighting given to seemingly contradictory pieces of evidence. I conclude with a discussion of how an alternative, inductive evaluation method and enhanced data elicitation may produce more reliable outcomes. In particular, I show how Verified's apparent effort to demonstrate compliance with the 'Guidelines for the use of language analysis in relation to questions of national origin in refugee cases' (Arends and Blommaert 2004) does not match the actual practice demonstrated by the company's reports.

2. LADO methods and debates

Following debates among academics and LADO practitioners, a set of 'Guidelines for the use of language analysis in relation to questions of national origin in refugee cases' was published in an effort to ensure the quality of language analysis reports (Arends and Blommaert 2004). The Guidelines call for reports to be authored by qualified linguists, to present their conclusions in qualitative terms relating to degree of certainty rather than in quantitative terms of percentage of certainty, to present verifiable data from the recording, to draw on published sources pertaining to the linguistic material that is presented, and to address the question of whether the applicant's speech is consistent with a named 'linguistic community' rather than with a place of origin, thereby acknowledging the possibility of structural variation and multiple influences on an individual's speech.

There has since been a perception that the discussion surrounding LADO practice and principles has been split into two 'camps': on the one hand, proponents of the Guidelines who argue for language analysis reports to be authored by qualified linguists and to assume the form of transparent scientific analyses (cf. Eades 2005; Fraser 2011; Patrick 2012), and on the other hand proponents of the opposing position which proposes that LADO casework should always have input from native speakers of the language(s) in question (cf. Cambier-Langeveld 2010, 2012; Nolan 2012). Eades (2005) presents a critique of contractors' qualification criteria for native speakers, which tend to emphasise the ability to apply the procedure set by the contractor itself as well as generic skills such as the ability to listen rather than linguistic training. Fraser (2011) and Patrick (2012) both contest that the agencies that carry out LADO typically lack expertise in linguistics and the languages under analysis and rely instead on the input of untrained native speaker analysts, which is difficult to verify. Verrips (2011) takes issue with the way in which conclusions are formulated by teams of native speakers and supervising experts who have no familiarity with the language that is being analysed, in particular with the way in which conclusions in such cases are based on subjective impressions rather than verifiable evidence.

On the opposing side of the debate, emphasis has been placed on particular skills of forensic analysis. In a short resolution passed in 2009[7] (see Moosmüller 2010), the International Association for Forensic Phonetics and Acoustics recognised that both linguists and trained native speakers who are not linguists can play a role in language analyses for the determination of origin, while specialised training is required from both to be able to perform a qualified analysis. Cambier-Langeveld (2010, 2012) reiterates this point and warns of the pitfalls of non-forensically trained linguists performing language analysis. She questions the assumption that non-native speaker linguists can carry out a language analysis without support from a native speaker, drawing parallels to descriptive linguistic fieldwork. Nolan (2012) emphasises the benefits of including native speakers' judgments in language analyses, arguing that they are more likely than trained linguists who are not native speakers to detect subtle phonetic and prosodic features, and therefore that native speakers can make a contribution to LADO that is complementary to that of trained linguists.

Criticism of existing LADO procedures has centred on two additional areas. The underlying assumption that language is a stable feature that can help establish a clear link between an individual and a place of origin has been questioned by researchers who point out that language is variable and that the use of variants is contextual (Reath 2004; Maryns 2004). Blommaert (2009) introduces a theoretical angle into this discussion, linking the critique to Vertovec's (2007) notion of superdiversity and arguing that globalisation and increased mobility can lead to more complex life histories, which in turn will be reflected in a more complex and differentiated linguistic repertoire (see also Jacquemet 2009; Spotti 2016). In a case study involving Kurdish, Maryns (2004) notes how the Belgian LADO procedure fails to recognise the migration history of an individual who had moved from Turkey to Iraq. In his analysis of a LADO report on Sudanese Arabic produced by the Dutch immigration authorities, Spotti (2016) argues that, in seeking to match a predefined structural concept of language with place, the analysts fail to take into consideration the indexical nature of linguistic variants and the potential hybridity in language that are an inevitable outcome of mobility. Spotti shows how the LADO report disregards style shifting and accommodation as well as the presence of features from the prestigious regional koiné that is spreading as a result of urbanisation. Addressing Arabic more generally, Rosenhouse (2013, 2017) notes that speakers are often multi-dialectal and that, alongside features that are shared or distinctive among neighbouring Arabic dialects, there are also 'new features' that are more volatile and which include regionally prestigious koiné-type forms (often derived from an urban dialect), features from Modern Standard Arabic or 'schooling effects', and loan words from contemporary contact languages. According to Patrick (2012), there is, for these and other reasons, an emerging consensus among researchers that LADO should address individuals'

socialisation history and their way of communicating in a particular setting (that of the interview) rather than place of origin.

Discussions have also drawn attention to issues surrounding the context and mode of data elicitation and assessment. While some studies point to the importance of focusing on (regionally) diagnostic features (cf. Simo Bobda, Wolf and Peter 1999; Rosenhouse 2013), others ask whether such focus might be anchored in an unrealistic notion of 'homogeneism' of language (cf. Eades and Arends 2004) and point to the difficulties of distinguishing between variation within a language variety and code-switching between speech varieties (cf. Maryns and Blommaert 2002). Taking a conversation analytical approach, Corcoran (2004) shows how Dutch immigration authorities' method of structuring the interview constrains the applicant's speech by controlling turn taking and topic, often resulting in very limited documentation of connected speech and conflicting interpretations about the goals of the interaction. She concludes (2004: 216) that it is difficult to compile an exhaustive list of diagnostic features since linguistic features are deployed creatively rather than in a way that presupposes a particular context.

In the following I will show how all three areas – issues surrounding analysts' qualifications and the input of native speakers, assumptions about the uniformity of language, and the mode of eliciting and contextualising data – remain issues of concern in the format of LADO reports adopted by the company Verified in recent UK asylum procedures. First, I review Verified's report structure and the way in which it contextualises data and deals with complex layers of features. I then introduce what I call an 'inductive' method of language analysis. This method assesses the plausibility of the applicant's claim to have been socialised in a particular location or locations by evaluating features of the applicant's speech in relation to known geographical and sociolinguistic isoglosses, looking for a meaningful or diagnostic co-occurrence of features. It relies both on documentation of connected speech, in the form of narration, and on targeted elicitation of a catalogue of phrases aiming to ensure comparability with verifiable control samples. In this way, it aims to address effectively issues of transparency and contextualisation of data. The method also offers a new approach to the collaboration setup between trained linguists and native speakers: It involves a trained linguist with expertise and active conversational knowledge of the languages that are being analysed, who draws for comparison on a comprehensive database of samples from a variety of locations, recorded, transcribed and processed by native speakers, which is publicly accessible and thereby verifiable.

3. Report structure and procedure

A number of studies (among them Corcoran 2004; Singler 2004; Maryns 2004; Blommaert 2009; Spotti 2016) offer a critical analysis of the way in which lin-

guistic data are assessed in individual LADO reports by different agencies, while Patrick (2016) discusses a corpus of Språkab reports, focusing on procedural rather than linguistic aspects. My aim is to present a corpus-based procedural and linguistic-methodological evaluation. I draw on a corpus of 50 reports by Verified (labelled 'LOID' – Linguistic Origin Identification) in cases of asylum seekers stating to be from Syria, 25 of them on Arabic and 25 on Kurmanji (Kurdish), delivered between 2015 and 2017, and on the audio recordings on which they were based. The corpus was constructed by the author, based on casework materials obtained with permission from the applicants' legal counsel. In all the cases the reports concluded that the data were inconsistent with the applicant's stated place of origin, and the Home Office relied on the reports for its decision to refuse the applicant's asylum application. In all cases considered here, the applicants appealed against that decision, supported by a counter-expertise report based on the original recording. In twelve cases altogether, a supplementary recording was carried out by the applicants' legal counsel, based on instructions provided by the counter-expert. In nine of the cases I received access to the outcome of the appeals, all but one of which have been successful. Some of the court decisions are available in the public domain as so-called 'reported' cases while others are not, but those too can be obtained from the courts by citing the appeal reference number. In order to ensure full protection of anonymity I have chosen not to disclose the appeal reference numbers.

Verified's reports follow an established routine seen in other LADO work (cf. Eades 2005; Patrick 2016): A recording of an interview with the applicant is assessed by a team consisting of a principal author, referred to as the 'linguist', and native speaker consultants, referred to in the reports as 'analysts'. The report is divided into sections on phonology, morphology, syntax and lexicon, addressing typically around five to six features each for phonology and morphology, fewer for syntax (usually a repetition of morphological features or features pertaining to grammatical lexicon and function words), and often two to three, and sometimes none at all, for lexicon. Examples are presented in the form of single words or sometimes combinations of two to three words, in a notation that is broadly oriented towards IPA but lacks morphological glossing. No indication is provided of the times in the recording from which examples are taken, and only three reports were accompanied by a complete transcription. Of those, one (for Arabic) used the Arabic script, thus leaving out most of the relevant phonological and phonetic detail, while the others used the orthography conventions for Kurmanji, capturing some variation in pronunciation but omitting important distinctions such as those between glottals and pharyngeals [h-ħ, ʔ-ʕ] and unvoiced and voiced velars [x-ɣ]. The discussion of each structural feature is accompanied by references to

academic publications. The report includes general remarks on method and biographical summaries of the contributors, identifying the 'linguist' by name and the 'analysts' by code number.

At a cursory glance, the format addresses some of the principles of the 'Guidelines for the use of language analysis in relation to questions of national origin in refugee cases' (Arends and Blommaert 2004; see also discussion in Eades 2009; Patrick 2012, 2016): The report purports to be authored by qualified linguists, it presents its conclusions in qualitative terms relating to degree of certainty, it presents verifiable data from the recording and draws on published sources, and it asks whether the applicant's speech is consistent with a named 'linguistic community', implicitly suggesting that wider issues of repertoire and socialisation are taken into consideration.

At closer scrutiny, however, the picture is different. The general remarks that accompany each report define 'linguist' as somebody who had 'academic studies in linguistics'. The corpus contains reports by altogether 14 authors, 2 of whom were responsible for almost half of the sample, having written 12 and 11 reports respectively. All authors have academic degrees, though in fact only one specialised in phonetics, while the others are graduates of programmes in languages, communication or cultural studies, six of them from universities in Romania that do not appear to have dedicated programmes in linguistics at all. None of the authors has any knowledge of Arabic or Kurdish. The 'analysts' are typically university graduates who studied in their home countries and left the region usually within a decade before the report date. None of them are linguistics graduates. This resembles the patterns observed in the work of other LADO agencies (cf. Eades 2009; Fraser 2009; Patrick 2012). Often the same person of Syrian background is listed as the 'analyst' for reports on both Arabic and Kurmanji, assessing the speech of applicants whose stated places of origin include Aleppo, Homs, Daraa, Damascus, Al-Hasaka, Ar-Raqqa and others.

The interview is conducted by phone or Skype. The interviewer is almost always one of the analysts and usually originates from Syria. Interviews are between 16 and 36 minutes long, though the typical length is 2123 minutes. The interviewer explains that the purpose of the recording is to assess the applicant's 'dialect' and instructs the applicant to speak for as long as possible in their home language about any topic of their choice without mentioning names or discussing their asylum claim. The interviewer then asks the applicant to name their place of origin and to provide a description of their home town or village, public buildings, landscape and surrounding towns and villages, their house, family members, job, education, as well as local dishes and festivities. The procedure is similar to that employed by other agencies (cf. Corcoran 2004; Eades 2005; Verrips 2010) except that the recording is always carried out remotely rather than face to face.

The reports make use of a more or less fixed batch of academic sources that tend to be replicated from case to case. For Arabic, the principal sources are Behnstedt's (1997) atlas of Syrian Arabic and Behnstedt and Woidich's (1985) atlas of Egyptian Arabic dialects as well as entries on various Arabic dialects from the *Encyclopedia of Arabic Language and Linguistics* (Versteegh 2012). It should be noted that few up-to-date published sources are available on the relevant dialects. In relation to general and historical phenomena as well as dialect classification, cursory reference is occasionally made to the works of Brustad (2000), Watson (2002), Owens (2006) and Versteegh (1997). For Kurmanji, the pool of references includes Mackenzie (1961), Öpengin and Haig (2014) and Thackston's (2006) online grammar of literary Kurmanji. None of these include data from Syria, and only Mackenzie (1961) relies on fieldwork in Iraq (see below), though his data on Kurmanji represent just six locations.

The pool of features that are taken into consideration follows the template of references and is thus largely the same from case to case for each respective language, too. For Arabic it includes under phonology the pronunciation of historical and Standard Arabic /q/, which may vary between [q, ʔ, g, ɢ]; Standard Arabic /ǧ/, which varies between [dʒ, ʒ, g, j]; Standard Arabic dental fricatives /ð, θ/, which may appear as [ð, θ], [d, t] or [z, s]; and the raising of the vowel /a/ to [ɛ] in final inflectional position. Reference is sometimes also made to the fronting or centralisation of /u/ to [ɪ, ə] and to the diphthongisation of historical and Standard Arabic /aw, ay/ to [oː, eː]. Under morphology the reports consider the presence or absence of a /h/ segment in pronominal inflection markers such as /hā/ 3.SG.F while syntax addresses unbound negation markers /mā, mū, mišš/ and the inflectional negation marker /iš, š/. For Kurmanji, the template largely follows Öpengin and Haig's (2014) description of the features of south-eastern Kurmanji (SEK) dialects in Iraq and neighbouring Hakkari province in Turkey: raising of /o/ to [u], fronting of /û/ to [iː] and simplification of /xw/ to [x] in phonology; and the form of the plural nominal attributive suffix /ê, ên, êd/, the oblique case marker /î/ in singular masculine nouns, and the 3SG subjunctive verb inflectional ending /it/ in morphology. Under syntax the reports occasionally list the presence of particular prepositions and the structure of possessive constructions.

Verified's principal innovation is the framing of the discussion as two juxtaposed 'hypotheses', each of which is tested in relation to its consistency with the data in the recording. The first hypothesis follows the applicant's stated place of origin in Syria. For Arabic, the reports compare data from the recording with published information on Arabic in Syria. For Kurmanji, in the absence of published descriptions from Syria, the reports rely on the input of the 'analysts' as well as on negative inference from documentation of Kurmanji attributed to Iraq

(see below), operating on the unrealistic assumption that the attestation of a form in Iraq necessarily implies that it is absent from Syria. The alternate hypothesis is always the same: For Arabic, the assumption is that the applicant is from Egypt and speaks the Cairo dialect, while for Kurmanji it is that the applicant speaks what Öpengin and Haig (2014) call 'southeastern Kurmanji' or SEK[8] (which the Home Office invariably interprets as an origin in Iraq). None of the reports provide a case-specific justification for the choice of alternate hypothesis. It might therefore be inferred that the choice of alternate hypothesis is one of convenience, allowing Verified to run an economy of scale in its reports and to replicate almost on a wholesale basis a template of cited references and diagnostic features.[9]

By framing the question in the form of two competing hypotheses, the reports take a deductive approach to the data. There are usually two 'analysts', one from Syria and the other from Egypt (for Arabic) or from Iraq (for Kurmanji). Each 'analyst' provides input in relation to just one of the hypotheses, identifying individual forms in the recording as consistent or inconsistent with that hypothesis. A decision is then taken on the overall degree of consistency with each of the hypotheses. The conclusion is presented in the form of a scale of seven points ranging from high certainty that the applicant's speech is consistent to high certainty that it is inconsistent with each respective hypothesis.

Table 1: Mapping of results by section and sub-section [K12170] author's adaptation of conclusions presented in the Verified report

Section	Feature	No. of examples cited	Consistency with 'eastern Al-Hasaka (Syria) variety', according to Verified	
			Inconsistent	Consistent
3.3.2	Kurmanji /o/	3	3	0
3.3.2	Kurmanji /u/	4	2	2
3.3.2	Kurmanji /xw/	3	1	2
3.3.2	phoneme /ʕ/	3	2	1
3.3.2	phoneme /ħ/	1	0	1
3.3.2	phoneme /-v-/	3	0	3
3.3.2	**Total**	**17**	**8**	**9**
3.3.2	**Overall result according to Verified**		**Inconsistent**	
3.3.3	plural izafe /-e/	2	0	2
3.3.3	plural izafe /-et/	2	0	2
3.3.3	oblique /-i/	2	2	0
3.3.3	3SG /e/	1	0	1
3.3.3	3SG /it/	2	2	0

3.3.3	singular f izafe /a/	2	0	2
3.3.3	singular f izafe /e/	2	2	0
3.3.3	singular m izafe /e/	1	0	1
3.3.3	singular m izafe /i/	2	2	0
3.3.3	**Total**	**16**	**8**	**8**
3.3.3	**Overall result according to Verified**		**Consistent**	
3.3.4	periphrastic marking	3	0	3
3.3.4	adpositions	2	0	2
3.3.4	enclitic /ʒi/	2	0	2
3.3.4	enclitic /iʃ/	2	2	0
3.3.4	**Total**	**9**	**2**	**7**
3.3.4	**Overall result according to Verified**		**Consistent**	
3.3.5	lexical items	3	3	0
	General result according to Verified		**'Most likely inconsistent'**	

What is not clear, and is not explained in the reports, is the method used to derive that conclusion from the data. Table 1 from case [K12170][10] presents the results in relation to the hypothesis that the applicant's origin is, as he states, in Qamishli in Syria (which Verified classifies as 'eastern Al-Hasaka variety' based on input from its 'analyst'). Even when taking Verified's interpretation of individual data examples at face value, a higher proportion of data seems to be consistent with the applicant's statement – both when considering the number of features (13/23, or 56%) and the number of example tokens listed (24/45, or 53%). Yet Verified's conclusion is that the data are 'most likely inconsistent' with the hypothesis. No hierarchy of features is presented that would suggest that more weight is given to some features than to others.

The inventory for the alternate hypothesis (SEK from Iraq) is almost a mirror image of the one presented in Table 1, drawing on the same features and the same examples but reversing the consistency assessment, except that some features have been removed from consideration altogether, while for features that are said to be consistent with the alternate hypothesis more example tokens were added. As a result, the consistency rate with the alternate hypothesis amounts to 10/17 features and 24/35 example tokens. The report arrives at the conclusion that the data are 'most likely consistent' with the alternate hypothesis. While I have not carried out a statistical analysis of the stated consistency rate of features across the entire corpus, the issue of the transparency of the weight given to individual forms is one of principle, as the method of arriving at an overall score is not addressed in any of the Verified reports examined.

4. Data contextualisation

Verified's interpretation of individual features and example tokens is, however, not always realistic. I shall illustrate this first with reference to Arabic, then to Kurmanji. Comparative dialect studies of Arabic (e.g. Jastrow and Fischer 1980; Versteegh 1997: Ch. 10) identify features that are diagnostic of particular regions, but there is also extensive discussion of diglossia and the effect of style shift and urbanisation on structural variation (cf. Walters 1996; Bassiouney 2009; Abu-Melhim 2014; Albirni 2016). Verified reports do not make any reference to variation and treat it instead as an 'inconsistency', which is usually interpreted as an indication that the speaker is deliberately manipulating their speech.

In [A07164] (stated origin in Aleppo) the report identifies both retention of Standard Arabic /a/ and a shift to /e/ in the feminine singular inflectional ending (so-called Imala), but argues that the use of /a/ supports an origin in Cairo. In fact, /e/ occurs in the recording in a great variety of words such as /mádrase/ 'school', /kibbe/ 'Kubbeh' and /sine/ 'year'. By contrast, /a/ appears in tokens that emulate formal speech, usually in phrases where there is style shift, as in the case of /madīna/ 'town' in /aqrab madīna ilayna/ 'the closest city to us', contrasting with colloquial /aqrab balad illna/, or which contain a uvular segment (a known constraint on the process known as Imala) as in /ḥadīqa/ 'municipal park/garden' and /míntaqa/ 'district'. Taking style and phonological conditioning into consideration, the alternation between the two forms is thus reconcilable with an origin in Aleppo, but not with an origin in Cairo, where only /a/ is found, in all cases, since the process known as Imala is not present in the Cairo dialect.

In [A05162] Verified refers to Behnstedt (1997: 1819) when claiming that the expected realisation of Standard Arabic /q/ in the applicant's stated place of origin, Ar-Raqqa, is [g], and infers that his occasional use of [ʔ] in this position alongside [g] is inconsistent and suggests an origin in Cairo. In reality, the variant [ʔ] is typical of urban speech, not just in Cairo but also across the Levant, including Syria. The use of an urban variant in the speech of a young man a generation or more after Behnstedt's data collection in rural settlements in the area is certainly plausible and well in line with observations on the spread of urban koinés in Arabic in general (cf. Albirni 2016: 183; Bassiouney 2009: 111114). Verified also interprets the occasional presence of other variants as distinctive of Cairo, although they are also features of Standard Arabic. Intriguingly, it attributes the sound [dʒ] used by the applicant, which is not found in Cairo Arabic, to the influence of Nile Delta Arabic, introducing an *ad hoc* modification to its alternate hypothesis. The Home Office refusal decision in this case, which relied on the Verified report, was overturned by the court, which adopted the argument of the

counter-expertise report that the Verified team lacked qualifications and that it failed to consider the possibility of style variation:

> The author of the [Verified] report is stated as [xx] who has no identifiable skills in the Arabic language of any dialect … the report relies on 2 hypotheses and there is no ability to consider Arabic in its wider context … the appellant's report provides reasoned conclusions and actively seeks to identify a wider variety of possibilities for the appellant's dialect.[11]

Patterns of local and free variation are not considered, either. In [K09170] (stated origin Al-Hasaka) Verified asserts with reference to Öpengin and Haig (2014) that the applicant's use of the form /dixazin/ 'you want' is inconsistent with Al-Hasaka, where instead of a simple phoneme /x/ the cluster /xw/ should be expected, thus /dixwazin/. The report's overall conclusion is somewhat long-winded: 'The language analysis somewhat suggests that the results obtained more likely than not are consistent with the linguistic community as stated in the [alternate] hypothesis' (SEK as spoken in Duhok province in Iraq and in south-eastern Turkey). On that basis the Home Office ruled that the applicant's likely place of origin was in Iraq. The contrast between /xw/ and /x/ is a permanent feature of Verified's template and is found in each and every report on Kurmanji, with reference to the same source. Öpengin and Haig's observations on SEK derive exclusively from the idiolect of the co-author Öpengin, who is in fact from Şemdinli in Turkey (cf. 2014: 147). On the contrast /x:xw/ they merely remark that 'the consonant group [xw] is simplified and delabialised to [x] in SEK' (2014: 157), citing just two examples: /xwê/ 'salt' and /xwîn/ 'blood'. Their comment is partly supported by Mackenzie (1961: 37) and Blau (1975: 229230), who imply that [xw] is somewhat rare in this group of dialects, though they only discuss very few tokens. However, none of these sources excludes the possibility that [x] can also occur in Syria. The Manchester Database of Kurdish Dialects (Matras et al. 2016) contains, at the time of writing, data on Kurmanji from some two-dozen locations in Syria and Iraq. A comparison suggests that the distribution of the variants /x/ and /xw/ does not map onto the country borders at all and that it may also vary across lexical items, possibly showing free variation, and that /x/ may even occur more frequently in Syria than in Iraq (Table 2). Indeed, for [K09170] Verified supplied its own transcription of the recording. In that transcription it documents its 'analyst', who is from Kobani in Syria, as saying both /xarin/ 'food' and /xwarin/ in the very same sentence. Verified's inference drawn from Öpengin and Haig's work in respect of Syrian Kurmanji is thus not only incorrect, but entirely baseless. A court appeal against the Home Office refusal decision was upheld in this case. The judge followed the counter-expertise position and stated in relation to Verified's LOID report: 'Its data was flawed, as was their reliance on published sources … the assumptions they made were unfounded and the

inferences drawn from the data unsafe', and further that there were 'errors and inconsistencies … which questions the reliability of Verified's linguistic analysis'.[12]

Table 2: Distribution of /xːxw/ in Manchester Database of Kurdish Dialects (selected locations with database sample numbers)

Sample code	Location	Country	'salt'	'food'
K021	Mosul	Iraq	xwê	xarin
K036	Duhok	Iraq	xê	xarin
K042	Zakho	Iraq	xwê	xwarin
K044	Akre	Iraq	xwê	xwarin
K039	Sersink	Iraq	xwê	xwarin
K033	Qamishly	Syria	xwê	xwarin
K037	Kobane	Syria	xwê	xarin
K043	Basselhâya	Syria	xo	xarin
K055	Derik (Al Hasakah)	Syria	xwea	xwarin
K100	Raqqa	Syria	xwê	xwarin
K101	Qesirdib	Syria	xwê	xwarin
K102	Jindires	Syria	xuwê	xarin
K103	Rajo	Syria	xê	xarin

The lack of linguistic expertise in the respective languages surfaces in the reports in instances where variables are discussed without the benefit of a diachronic appreciation of the structures concerned. One of Verified's permanent template features for Kurmanji is the alternation of /û/ and /î/. This is the outcome of a process of fronting of the historical long vowel [uː]. In [K11163] the report cites the form /biçin/ 'they go' as an example for such fronting and concludes that the applicant's speech is inconsistent with Al-Hasaka but consistent with SEK. In fact, while the infinitive form is /çûn/ 'to go', the inflected forms /biçin/ (subjunctive) and /diçin/ (indicative), with a short /i/, appear in all Kurmanji varieties, a process that is quite distinct from the regionally distinctive fronting of [uː] to [iː], while fronting never appears in the infinitive /çûn/ 'to go' in any known dialect. This makes clear the risks of undertaking an analysis without direct access to historical-linguistic and dialectological expertise in the language concerned.

5. Complex patterns

Most of the cases in the corpus involve young males between the ages of 17 and 25; many were interviewed by the Home Office upwards of two years after leaving their initial place of socialisation. In most cases they report having spent prolonged periods of time in the company of other migrants from other countries. Some report to have flatmates from other countries who speak other varieties

of Arabic and Kurdish, respectively. In at least five cases[13] applicants reported to have a parent who had been a migrant, and in an additional five cases Arabic-speaking applicants reported to have left Syria at the age of 14 or younger and to have lived for several years in Egypt before migrating to Europe. None of the reports take these life histories into consideration in the assessment of the data.

In [K03170], the applicant reports that his family had moved from Derik (on the Iraqi border) to Qamishli (in northern-central Syria, bordering on Nusaybin in Turkey). Verified's first hypothesis does not take the possibility of dialect mixing into account and assumes instead a variety that is consistent with Qamishli. It then attributes features that are consistent with Derik to (neighbouring) SEK and concludes on that basis that the client is not likely to belong to the linguistic community of Qamishli.

In [A05161] (stated origin Aleppo), the recording contains a mixture of features of Egyptian and Syrian Arabic, such as word stress in /mádrasa/ 'school' (Syrian) alongside /madrása/ (Egyptian), which was not noted by Verified (whose reports generally do not pay attention to prosody), nominal negation with /mū/ (Syrian) alongside /mišš/ (Egyptian), and a few instances of [g] (Egyptian) alongside a majority of tokens in [dʒ] (Syrian). This seems consistent with the applicant's narrative according to which he left Syria and moved to Egypt at the age of 13. Verified notes the Syrian features but favours the hypothesis that the applicant is from Egypt; constrained by the structure of its own report template, Verified seems to give itself no choice but to favour one self-contained hypothesis over another. Verified's view and the ensuing Home Office refusal decision was initially supported on appeal by the First Tier Tribunal but then dismissed upon further appeal by the Upper Tribunal, which stated, with reference to Verified's report:

> The report failed to take into account the appellant's period of residence in Egypt, or to explain why this could not account for the features of 'Cairo Arabic' present in his speech. Nor did the report explain the presence of aspects of Syrian Arabic which the report acknowledges appeared within his speech.[14]

[A10160] is a further case that shows a mixture of dialectal forms that can be attributed respectively to the Syrian and Egyptian dialects of Arabic. The applicant reports to have left Daraa in Syria at the age of 16. His very first utterance in the interview is /š-lōnak/ 'how are you?', which is a distinctively Syrian greeting and might be regarded as having an emblematic function in the context of flagging competence in Syrian Arabic. His speech then shows several instances of alternation between Syrian and Egyptian forms, as in /hēk/ 'like this' (Syrian) alongside /kida/ (Egyptian); /ayy šey/ 'anything' (Syrian) alongside /ayy ḥāga/

(Egyptian). He also shows frequent single-word insertions from Standard Arabic, as in /daǧāǧ/ 'chicken' (colloquial /ǧāǧ/) and /ṣaġīra/ 'small' (colloquial /zġīra/), which might be interpreted as an effort to avoid dialectal forms, or possibly as an effort to accommodate to a formal setting. Uncommented on by Verified is his use of distinctively Egyptian word stress, as in /mudarrísa/ 'teacher' and /muqābála/ 'interview'. He consistently uses the distinctively Syrian negation marker /mū/ in nominal predications but shows some instances of hypercorrection where he extends that marker to finite verbs, as in /mū baʕref/ 'I don't know' and /mū waṣalt/ 'I did not arrive', where both Syrian and Egyptian Arabic in fact use the same negation marker /mā/.

Another case of dialect mixture is displayed in [A04162]. The applicant is 17 years old and reports to have grown up in a village near Damascus but to have left Syria at the age of 13 and then lived in Egypt for two years. He further reports that he shares a flat with an Egyptian person. The applicant says /təgīb/ 'you bring', using the distinctively Egyptian-Cairene [g], but immediately self-repairs to (Syrian) /təǧīb/. He shows phonological hypercorrections in /inǧilīzī/ 'English', where he substitutes [g] in a loanword by what he apparently perceives to be the correct Syrian counterpart [dʒ], and in /mintaqá/ 'area', apparently aiming to diverge from Egyptian /mintáqa/, while Syrian has /míntaqa/; and morphological hypercorrection in /mū fhimət/ 'I didn't understand' and /mū štaġalət/ 'I didn't work', over-generalising, as in the case above, the nominal negation marker /mū/ to finite verbs.

Of the 25 Arabic recordings in the corpus, I found evidence of distinctively Egyptian forms in 11 cases. In cases where a life-history narrative is not accessible, an assessment is required as to which of the two layers, Syrian or Egyptian, is more likely to represent the substrate. Following Moosmüller (2011: 185), possible clues might lie in features that could be identified as 'salient' and could be assessed in relation to the speaker's apparent level of co-operation (though detailed research and what might be perceived by speakers of Arabic as 'salient' is still lacking). It might be argued, for instance, that the presence of Egyptian word-level stress reveals a layer that is more difficult to suppress in speech production, while the emblematic greeting and the hypercorrections might strengthen the view that the applicant is making an effort to emulate a Syrian dialect. But when a speaker does recount a history that spans different countries or close encounters with speakers from different countries (such as sharing residence with Egyptians over a period of time at a makeshift refugee camp in Calais or an allocated flat in the UK), we do not know whether he is 'manipulating' his speech. It is equally possible that during the interview he is merely guided by the assumption that he is expected to use more Syrian than Egyptian forms and is therefore aiming to supress forms that might appear to be Egyptian. This might lead him to produce

forms that do not actually exist in the normal inventories of either of the two varieties. If this can be evidenced, then such an effort may well point to insecurity arising through the setting and the stakes that it entails rather than a deliberate attempt to mislead.

6. Toward an inductive-dialectological approach

In this final section I briefly introduce a method that addresses some of the problems alluded to in the review of the corpus of Verified reports. Taken for granted is firstly that the language analysis is carried out by a trained linguist with thorough knowledge of the languages that are the subject of the analysis, here (Levantine and Egyptian) Arabic and Kurmanji Kurdish. I draw on the example of [A06171] (stated origin in Daraa), where the method was used in a counter-analysis of the Verified report. The method aims to address challenges of data elicitation, contextualisation and verifiability.

In the majority of cases in the corpus, the questions asked during the interview produce 'descriptions' rather than 'reports' or 'narrations' (cf. Rehbein 1984). Structurally such responses are characterised by the chaining of noun phrases, often embedded into a series of existential predications, or at best a listing of routines that are expressed through serialisation of verbs in the same tense (usually habitual past) and often with the same subject/agent (usually 1SG). This limits the amount of grammatical and especially inflectional material that is available for analysis (cf. Corcoran 2004). Inspired by Labov's (1972) narrative interview technique, Singler (2004) proposes a sociolinguistic interview for language analyses that aims to engage the speaker and provoke a more intense flow of speech. Table 3 presents some of the questions that were used to guide supplementary data elicitation for a counter-analysis. They are designed to prompt narration in the sense of an organised series of utterances that reconstruct a sequence of real-life events. In order to ensure the linguist's impartiality, the recording was carried out by a speaker of Arabic who was appointed by the applicant's legal representative and provided with written instructions to read out the questions in Standard Arabic but to ask the applicant to answer them in his 'vernacular' variety (identifiable to most Arabs by the term *ʕāmiyya*).

Table 3: Examples of questions used to prompt narration

Do you know any films or television series? Describe one of your favourite episodes
Tell me a story that your grandmother, or another elderly relative, or acquaintance, told you, or that you read or heard at school; it can be a fairy tale, or something about history, or somebody's personal story (something that happened to them)
Can you remember an incident in your life when you were scared? Describe to me what happened

Table 4: Sample of feature presentation (original includes times)

The speaker shows fronting or centralisation of historical short /u/ to /i/ [ɪ, ə], sometimes in variation with /u/: *kull, kill* 'all' *kint* 'I was' *gimt* 'I got up' *miškile* 'problem' *kinna* 'we were' *hiwwe* 'he' This development is attested in the Arabic dialects of Syria and south-eastern Turkey, and in some dialects of Lebanon, as well as in parts of the Gulf regions

Table 5: Examples of isogloss-based assessment of features

a. The alternation of /ǧ/ or /ž/ for historical /ǧ/ is consistent with northern Levantine Arabic, broadly speaking north of Jerusalem, including Syrian Arabic, and may also occur in other individual locations.
b. The fronting of historical /u/ to /i/ is consistent with Arabic dialects of Syria, Lebanon and south-eastern Turkey, as well as with some dialects of Mesopotamia and the Gulf regions
c. The use of negation particles /mū/ in nominal predications and /mā/ unaccompanied by a suffix with inflected verbs and possessive constructions is consistent with Syrian Arabic as well as Mesopotamian and Gulf Arabic.
d. The use of /g/ with occasional variants /ʔ/ and /q/ for historical /q/ is consistent with the Arabic varieties of eastern and southern Syria and northern Jordan, and possibly also with the varieties of Upper Egypt, assuming exposure to urban varieties or urban-based regional koinés.

Verrips (2010) introduces a language analysis template that avoids a pre-set hypothesis and relies instead on an integrated assessment of the inventory of forms from the recording. Verrips's method sets out the categorisation of forms by structure-based themes but does not prescribe the way in which the report's conclusions are to be presented. The method adopted here follows a similar principle, listing examples from the recording under the headings phonology, morphology, syntax (where relevant) and lexicon. But the findings are presented consistently in relation to isoglosses (that represent the geographical distribution of the respective features in the Arabic-speaking regions), first individually, with examples from the recording (Table 4), and then as a summary overview (Table 5) that offers a broader contextualisation of each of the features.

An inductive dialectological assessment was applied both to the Home Office recording carried out by Verified and to the supplementary recording carried out using interview questions such as those in Table 3, with identical results. The report's conclusion is then formulated in relation to the intersection of isoglosses, giving a holistic perspective on the data that derives inductively from that contextualisation rather than in relation to a pre-set hypothesis:

> All features used by the client in the recording are consistent with the dialects of southern Syria to which that of Daraa belongs, assuming some exposure to urban-based regional koinés. In the particular combination in which they occur in the client's speech, the features used by the client are consistent exclusively with the Arabic spoken in southern Syria under the influence of urban varieties. The linguistic evidence therefore supports the client's narrative according to which he is from Daraa in Syria.

In addition to the re-assessment of the Home Office recording, and the assessment of a supplementary recording that attempted more systematically to elicit narrative speech (using the question template depicted in Table 3), a third component was included in the analysis. This component relies on the elicitation of phrases that are designed to capture key diagnostic features of Arabic dialects in the areas of lexicon, lexico-phonology and morpho-syntax. The phrases were taken from a comprehensive template of altogether 1,065 items, including single words, verb inflection paradigms, and short sentences. The template served as a questionnaire designed to capture morpho-syntactic and lexico-phonological variation among Arabic dialects as part of a research project carried out between 2012 and 2017. The questionnaire was administered to bilingual speakers using English as the elicitation language; data were recorded, transcribed and entered into an online Database of Arabic Dialects (Matras et al. 2017)[15] At the time of writing the database contains material from over 20 locations in 15 different countries, including samples from both Daraa and Cairo, which were drawn upon for comparison with speech elicited from the applicant.

In order to obtain data for comparison from the applicant, a shortened version of the database questionnaire was designed. It was administered to the applicant in Standard Arabic by a third party, in order to maintain the linguist's impartiality. Some phrases, which showed traces of accommodation through direct replication of the Standard Arabic prompt, were eliminated from consideration. Altogether around one hundred elicitation phrases were analysed, using both the above method of isogloss comparison and, in addition, specific comparison with the database samples for Daraa and Cairo (Table 6).

Table 6: Extract from questionnaire comparison

English phrase	Applicant	SYR002 Daraa	EGY001 Faiyum Oasis/ Cairo
here	hōn	hōn	henā
hungry	ǧūʕān	ǧūʕān	gaʕān
don't go	lā trūḥ	lā trūḥ	mā tiruḥ
two sons and a daughter	waladēn ū binit	waladēn ū binit	waladīn wa bint

I do not want to talk	mā biddī ʔaḥčī	mā biddī ʔaḥčī	miš ʕawz ʔatkallam
he is not in the desert	hiwwe mū bi ṣ-ṣaḥra	hiwwe mū bi ṣ-ṣaḥra	howa miš fi ṣ-ṣaḥara
I do not have a sister	mā ʕindī ʔuxit	mā ʕindī ʔuxit	mā ʕandī-š ʔuxt
this car is ours	hāy is-sayyāra la-ʔilnā	hāy is-sayyāra ʔilnā	el-ʕarabiya dī bitaʕtnā
after five minutes he started to talk	baʕd xamas dagāyig balaš bi-l-ḥačī	baʕd xamas dagāyig hiwwe balaš yiḥčī	baʕd xamas daʔyeʔ bada yitakallam

The comparison with the database samples (Table 6) provides very strong independent support that the applicant's speech is consistent with that of Daraa. Yet in this very same case, Verified concluded – arguing on the basis of urban variants such as /ʔ/ and the retention of /h/ in the pronunciation of pronominal inflection (alongside its frequent omission by the applicant) – that the results 'somewhat suggest' inconsistency with the linguistic community of Daraa. In the second part of its report, despite listing an equal number of features (five each) as being 'consistent' and 'inconsistent' with the speech of Cairo, it concluded that the results 'somewhat suggest' consistency with Cairo. Citing the Verified report, the Home Office refused the applicant's claim for asylum. The case went to appeal, citing the evidence listed above in Tables 4, 5 and 6. The judge of the First-Tier Immigration Tribunal overturned the Home Office decision and followed the arguments of the counter-expertise report. He stated:

> Verified have not mentioned a single item from the appellant's speech that is consistent with Egyptian Arabic but not consistent with another Arabic dialect … It seems to me that the whole process of picking a 'hypothesis 2' to be juxtaposed against the appellant's claimed linguistic community of origin is at least questionable … I prefer [the] reference to a co-occurrence in the appellant's speech being sufficient to confirm the appellant's origin in Syria … I also accept that criticisms […] relating to the Linguist used by Verified in this case are valid. In particular, the Linguist is not said to have any relevant qualifications in the Arabic language … I also accept that the reliability of [the] evidence is strengthened by a database of recordings of Arabic speakers maintained by Manchester University.[16]

The judge thus accepted the merits of the inductive dialectological analysis over the juxtaposition of two hypotheses, the need for first-hand expertise in the language that is being analysed, and the benefits of drawing on a comparison with a publicly available and verifiable control corpus of dialect data.

7. Conclusion

Campbell (2013: 685656) concludes his discussion of UK language analysis procedures with the bold statement that 'the policy of language analysis is fundamentally political' and that in LADO reports 'the language of science operates as an illusion of smoke and mirrors to obfuscate flawed assumptions about language use'. This echoes Blommaert's (2009: 421) critique of UK LADO procedures as driven by an ideology of language that makes 'time and space static' and disregards the sociolinguistic reality of language as a dynamic repertoire of resources. Verified's entrance into the LADO stage in the UK appeared to signal a change in the quality of language analysis reports: The company introduced a report template that cited a large number of examples (usually between 30 and 50 in each of its two report sections) compared to the relative paucity of examples used by its predecessor contractor Språkab (typically around 10 altogether), it makes use of extensive citation of published references, and it introduced a method that purports to be scientifically transparent, by which two hypotheses are juxtaposed. However, the comparative corpus discussion presented above shows that the format and presentation of Verified's reports do not alleviate the concerns addressed in the research literature. The framing of the Verified analysis as a juxtaposition of two hypotheses to be proven or disproven, and the dense use of academic references, serve as a way of presenting the reports as scientific. But they also replicate the 'monoglot ideology' and notions of 'homogeneism' by excluding the possibility of variation, layering and indexicality of linguistic forms (as defined, for instance, by Silverstein 2003).

The key flaw is the absence in the reports of any explicitly formulated holistic evaluation of the data that would contextualise the applicant's speech within a realistic understanding of isogloss intersection (both spatial and social) while giving consideration to communicative strategies and reported life-history trajectories. This shortcoming is inherently part of the process whereby native speaker 'analysts' are asked for their subjective, binary judgment on individual forms, while the 'linguist' appears to lack either the qualifications or a transparent method to integrate their input into a reasonably balanced conclusion. This approach precludes any outcomes that are not pre-formulated as one of the two hypotheses. This also means that Verified's formulation that 'the data are (in)consistent with the linguistic community of X' can appear to be compliant with the guidelines alluded to in the introductory remarks (cf. Arends and Blommaert 2004), by allowing the possibility of a stratified socialisation, yet the Home Office invariably interprets such a statement as pertaining to the applicant's place of origin.

Moreover, Verified's choice of alternate hypothesis is wholesale and appears to be motivated by criteria that are external to the individual case. For Arabic these

include, on the political and administrative side, the realisation that some applicants stating to be from Syria have either travelled via Egypt or originate from Egypt. Notwithstanding the possibility that this may be true in some cases, along with the procedural flaws, this predetermined hypothesis runs the risk of biasing the outcome of the analysis. For Kurmanji, the alternate hypothesis, which is invariably interpreted by the Home Office as suggesting an origin in Iraq, seems merely to exploit the opportunity offered by a single, easily accessible online publication (Öpengin and Haig 2014) in which data on the so-called SEK variety are presented based on the idiolect of a single individual who originates from Turkey and not from Iraq. For the contractor, the method allows an economy of scale while offering a procedure that is seemingly more robust than the one applied by its predecessor Språkab.

The review of court decisions in individual cases, however, shows that the use of counter-analysis reports to support an appeal is helpful and indeed essential in bringing to the fore doubts about the expertise of the Verified teams and critique of their failure to take account of variation and repertoire complexity. It seems that some of the concerns expressed in research on LADO are resonating within the legal system, possibly counteracting political pressures: The court decisions cited above, and others, emphasise the need to take into consideration applicants' history of socialisation in different places, where applicable, and attribute importance to clearly demonstrating the linguistic expertise of the report authors, and to the verifiability of the input of native speakers.

It therefore seems that much can be achieved by applying alternative methods of data elicitation and assessment more systematically, as proposed already by Verrips (2010), Patrick (2012, 2016) and others. The online Manchester databases for Arabic and Kurdish dialects have introduced a major new factor into the evaluation of LADO cases from Syria, allowing more accurate contextualisation at least of so-called 'static' (spatially bound) features.[17] This underlines the importance of so-called 'baseline' research in descriptive dialect geography, which is often de-prioritised by UK research funding schemes in favour of ventures that put forward agendas that are more theoretical. In addition to a broader descriptive basis, more use should be made in language analysis of techniques that seek to elicit connected speech and which give consideration to the structure of narration. More weight should also be given to the discourse-analytical dimensions of the interview, insisting that turn-taking management, accommodation, self-repairs and repetitions and prosodic features are properly analysed.[18]

In conclusion, in this article I have argued for a new protocol for use in LADO analysis: It relies on a qualified linguist with first-hand expertise in the descriptive historical-linguistic and dialectological study of the language in question leading the analysis, with reference to a verifiable corpus of comparative data compiled from and with the help of native speakers. It requires a sample of speech from

the applicant that is designed to elicit connected narration, on the one hand, and targeted structures that can be compared to a control corpus, on the other. And it relies on an inductive approach or an 'open hypothesis' (e.g., 'the applicant speaks any possible dialect of Arabic') rather than a fixed, pre-set alternate hypothesis, and on the assessment of the co-occurrence of features and the extent to which they render the applicant's statement about their own life history and socialisation plausible or not.

About the author

Yaron Matras is Professor of Linguistics at the University of Manchester and founder of the Multilingual Manchester research unit. He is the author of *Language Contact* (CUP, 2009), *Romani in Britain: The Afterlife of a Language* (EUP, 2009), *A Grammar of Domari* (Mouton, 2012), *Romani: A Linguistic Introduction* (CUP, 2002)

Notes

1. http://www.bbc.co.uk/news/uk-england-35439030
2. http://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN06805#fullreport
3. https://www.gov.uk/government/publications/immigration-statistics-october-to-december-2016/asylum
4. https://www.gov.uk/government/speeches/language-analysis-testing-authorisation-2013-palestinian-syrian-kuwaiti-no-2
5. *Secretary of State for Home Department v. MN and KY*, UKSC 30 [2014].
6. UK Home Office reply to the author's Freedom of Information request; FOI reference 48051, 24 April 2018.
7. http://www.iafpa.net/the-association/resolutions/
8. There is one exception, discussed below, where the alternate hypothesis relates to Sorani Kurdish as spoken in Iraq.
9. The Sorani template referred to in the previous footnote is used as an alternate hypothesis in cases of Sorani-speaking applicants stating they are from Iran.
10. I use a sample-internal referencing system; each case is cross-referenced to the Home Office case number and to the contractor's original report number and date. Codes in K- indicate reports on Kurmanji, those in A- on Arabic.
11. First Tier Tribunal (appeal no. with author), August 2016, p. 6.
12. First Tier Tribunal (appeal no. with author), October 2017, p. 89.
13. I have only had access to some of the case files containing details of applicants' self-declared life history.
14. Upper Tribunal (appeal no. with author), May 2017, p. 10.
15. http://www.arabic.humanities.manchester.ac.uk/database-of-arabic-dialects/

16. First-Tier Immigration Tribunal (appeal no. with author), March 2018, pp. 14–16.
17. It should be noted that most of the data for the Kurdish database was collected in situ in the Middle East, while for the Arabic database some were collected in the Middle East while others in the UK.
18. To that end, the Home Office and its contractors should be obliged to provide full transcriptions of the interviews.

References

Abu-Melhim, A. (2014) Intra-lingual code alternation in Arabic: the conversational impact of diglossia. *Theory and Practice in Language Studies* 4(5): 891902. https://doi.org/10.4304/tpls.4.5.891-902

Albirni, A. (2016) *Modern Arabic Sociolinguistics*. London: Routledge.

Arends, J. and Blommaert, J. (2004). Guidelines for the use of language analysis in relation to questions of national origin in refugee cases. *International Journal of Speech, Language and the Law,* 11(2): 261266. https://doi.org/10.1558/sll.2004.11.2.261

Bassiouney, R. (2009) *Arabic Sociolinguistics.* Edinburgh: Edinburgh University Press. https://doi.org/10.3366/edinburgh/9780748623730.001.0001

Behnstedt, P. (1997) *Sprachatlas von Syrien* vol. 1: *Kartenband und Beiheft.* Wiesbaden: Harrassowitz.

Behnstedt, P. and Woidich, M. (1985) *Die ägytisch-arabischen Dialekte.* Tübingen: Reichert.

Blau, J. (1975) *Le kurde de Amadiya et de Djabal Sindjar*. Paris: Klincksieck.

Blommaert, J. (2009) Language, asylum and the national order. *Current Anthropology* 50(4): 415441. https://doi.org/10.1086/600131

Brustad, K. B. (2000) *The Syntax of Spoken Arabic.* Washington, DC: Georgetown University Press.

Cambier-Langeveld, T. (2010) The role of linguists and native speakers in language analysis for the determination of speaker origin. *International Journal of Speech, Language and the Law* 17(1): 67–93. https://doi.org/10.1558/ijsll.v17i1.67

Cambier-Langeveld, T. (2012) Clarification of the issues in language analysis: a rejoinder to Fraser and Verrips. *International Journal of Speech, Language and the Law* 19(1): 95108. https://doi.org/10.1558/ijsll.v19i1.95

Campbell, J. (2013) Language analysis in the United Kingdom's Refugee Status Determination system: seeing through policy claims about 'expert knowledge'. *Ethnic and Racial Studies* 36(4): 670690. https://doi.org/10.1080/01419870.2011.634506

Corcoran, C. (2004) A critical examination of the use of language analysis interviews in asylum proceedings: a case study of a West African seeking asylum in the Netherlands. *International Journal of Speech, Language and the Law* 11(2): 200221. https://doi.org/10.1558/sll.2004.11.2.200

Craig, S. (2012). The use of language analysis in asylum decision-making in the UK - a discussion. *Journal of Immigration Asylum and Nationality Law* 26(3): 255268.

Eades, D. (2005) Applied linguistics and language analysis in asylum seeker cases. *Applied Linguistics* 26(4): 503526. https://doi.org/10.1093/applin/ami021

Eades, D. (2009) Testing the claims of asylum seekers: the role of language analysis. *Language Assessment Quarterly* 6(1): 3040. https://doi.org/10.1080/15434300802606523

Eades, D and Arends, J. (2004) Using language in the determination of national origin of asylum seekers: an introduction. *International Journal of Speech, Language and the Law* 11(2): 179199. https://doi.org/10.1558/sll.2004.11.2.179

Fraser, H. (2009) The role of 'educated native speakers' in providing language analysis for the determination of the origin of asylum seekers. *International Journal of Speech, Language and the Law* 16(1): 113138. https://doi.org/10.1558/ijsll.v16i1.113

Fraser, H. (2011) The role of linguists and native speakers in language analysis for the determination of speaker origin: a response to Tina Cambier-Langeveld. *International Journal of Speech, Language and the Law* 18(1): 121130. https://doi.org/10.1558/ijsll.v18i1.121

Home Office (2017) Language analysis. Version 20.0. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/591701/Language-analysis-v20.pdf

Jacquemet, M. (2009) Transcribing refugees: the entextualization of asylum seekers' hearings in a transidiomatic environment. *Text & Talk* 29(5): 525546. https://doi.org/10.1515/TEXT.2009.028

Jastrow, O. and Fischer, W. (eds) (1980) *Handbuch der arabischen Dialekte.* Wiesbaden: Harrassowitz.

Labov, W. (1972) *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

MacKenzie, D. N. (1961) *Kurdish Dialect Studies*. London: Clarendon Press.

Maryns, K. and Blommaert J. (2002) Pretextuality and pretextual gaps: on de/refining linguistic inequality. *Pragmatics* 12(1): 1130. https://doi.org/10.1075/prag.12.1.02mar

Maryns, K. (2004) Identifying the asylum speaker: reflections on the pitfalls of language analysis in the determination of national origin. *International Journal of Speech, Language and the Law* 11(2): 240260. https://doi.org/10.1558/sll.2004.11.2.240

Matras, Y. et al. (2016) The dialects of Kurdish web resource. The University of Manchester. http://kurdish.humanities.manchester.ac.uk/

Matras, Y. et al. (2017) Database of Arabic dialects. The University of Manchester. http://www.arabic.humanities.manchester.ac.uk/database-of-arabic-dialects/

Moosmüller, S. (2010) IAFPA position on language analysis in asylum procedures. In K. Zwaan, M. Verrips and P. Muysken (eds) *Language and Origin The Role of Language in European Asylum Procedures: Linguistic and Legal Perspectives* 4347. Nijmegen: Wolf Legal Publishers.

Moosmüller, S. (2011) Evaluation of the speech behaviour of reference speakers. *International Journal of Speech, Language and the Law* 18(1): 179205.

Nolan, F. (2012) Degrees of freedom in speech production: an argument for native speakers in LADO. *International Journal of Speech, Language and the Law* 19(2): 263289. https://doi.org/10.1558/ijsll.v19i2.263

Öpengin, E. and Haig, G. (2014) Regional variation in Kurmanji: a preliminary classification of dialects. *Kurdish Studies* 2(2): 143176.

Owens, J. (2006) *A Linguistic History of Arabic*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199290826.001.0001

Patrick, P. (2012) Language analysis for determination of origin: objective evidence for refugee status determination. In L. M. Solan and P. M. Tiersma (eds) *The Oxford Handbook of Language and Law* 533546. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199572120.013.0039

Patrick, P. (2016) The impact of sociolinguistics on refugee status determination. In R. Lawson and D. Sayers (eds) *Sociolinguistic Research: Application and Impact* 235–256. London: Routledge.

Reath, A. (2004) Language analysis in the context of the asylum process: procedures, validity, and consequences. *Language Assessment Quarterly* 1(4): 209233. https://doi.org/10.1207/s15434311laq0104_2

Rehbein, J. (1984) Beschreiben, Berichten und Erzählen. In K. Ehlich (ed.) *Erzählen in der Schule* 67124. Tübingen: Narr.

Rosenhouse, J. (2013) General and local issues in forensic linguistics: Arabic as a case study. *Comparative Legilinguistics* 15: 5367. https://doi.org/10.14746/cl.2013.15.4

Rosenhouse, J. (2017) A forensic linguistic problem: asylum seekers' dialect identification difficulties in under-documented adjacent Arabic dialects. *International Journal of Legal Discourse* 2(1): 113123. https://doi.org/10.1515/ijld-2017-0005

Silverstein, M. (2003) Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23(34): 193229. https://doi.org/10.1016/S0271-5309(03)00013-2

Simo Bobda, A., Wolf, H-G. and Peter, L. (1999) Identifying regional and national origin of English-speaking Africans seeking asylum in Germany. *Forensic Linguistics* 6(2): 300–319. https://doi.org/10.1558/sll.1999.6.2.300

Singler, J. V. (2004) The 'linguistic' asylum interview and the linguist's evaluation of it, with special reference to applicants for Liberian political asylum in Switzerland. *International Journal of Speech, Language and the Law* 11(2): 222239. https://doi.org/10.1558/sll.2004.11.2.222

Spotti, M. (2016) Sociolinguistic shibboleths at the institutional gates: language, origin and the construction of asylum seekers' identities. In K. Arnaut, J. Blommaert, B. Rampton and M. Spotti (eds) *Language and Superdiversity* 261278. London: Routledge.

Thackston, W. M. (2006) *Kurmanji Kurdish: A Reference Grammar with Selected Readings*. [online version]

Verrips, M. (2010) Language analysis and contra-expertise in the Dutch asylum procedure. *International Journal of Speech, Language and the Law* 17(2): 279294.

Verrips, M. (2011) LADO and the pressure to draw strong conclusions: a response to Tina Cambier-Langeveld. *International Journal of Speech, Language and the Law* 18(1): 131143. https://doi.org/10.1558/ijsll.v18i1.131

Versteegh, K. (1997) *The Arabic Language*. Edinburgh: Edinburgh University Press.

Versteegh, K. (ed.) (2012). *Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill.

Vertovec, S. (2007) Super-diversity and its implications. *Ethnic and Racial Studies* 30(6): 10241054. https://doi.org/10.1080/01419870701599465

Walters, K. (1996) Diglossia, linguistic variation, and language change in Arabic. In M. Eid (ed.) *Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics.* Vol. 8, 157200. Amsterdam: John Benjamins.

Watson, J. (2002) *The Phonology and Morphology of Arabic*. New York: Oxford University Press.