

# SPH-LENS: A Socio-Political-Historical Framework for Early Warning of Arabic Language Attrition as a National Security Threat

Mostafa Ahmed, Dr. Mohamed Shadi

Al Habtoor Research Centre – Dubai

DOI: <https://dx.doi.org/10.47772/IJRISS.2025.91100287>

Received: 10 November 2025; Accepted: 20 November 2025; Published: 06 December 2025

## ABSTRACT

Arabic is widely celebrated as a “living and powerful language” – boasting over 400 million speakers and official status across 22 Arab states – yet evidence of its declining societal role reveals an underappreciated security risk. This paper reframes the erosion of Arabic not as a cultural footnote but as a first-order national security challenge. Building on the Socio-Political-Historical (SPH) conceptual foundation, the paper integrates sociolinguistic capital theory and linguistic hegemony to illustrate how Arabic language attrition – manifesting through domain losses and prestige decline – can unravel social cohesion. The paper critique conventional language vitality indices (e.g., UNESCO’s World Atlas of Languages and Ethnologue’s EGIDS) for their static, lagging indicators that deem Arabic “Safe” and argue they fail to capture dynamic early-warning signs of shift. In response, the paper present SPH-LENS (Socio-Political-Historical Language Early Warning and National-Security System), a new framework that operationalises Arabic attrition risk monitoring. **SPH-LENS** comprises three core dimensions – Socioeconomic (**S**), Political (**P**), and Historical (**H**) – each justified theoretically and populated with concrete, measurable indicators from existing global data (e.g., UIS, OpenAlex, WIPO, W3Techs, Wikimedia, EF EPI, Constitue Project, Freedom House, World Bank GTMI, and UNESCO media statistics).

By transparently synthesising these indicators into a composite risk index, **SPH-LENS** provides governments an evidence-based dashboard to detect early attrition “red flags” before intergenerational transmission collapses. Through examples in Tunisia, the UAE, Morocco, and Egypt, the paper show how **SPH-LENS** would flag domain shifts (such as education and digital media pivoting to English/French) that foreshadow broader linguistic displacement. The framework’s policy relevance is emphasised: it offers a replicable tool for strategic intervention, treating linguistic vitality as an essential pillar of national security. The paper concludes by positioning **SPH-LENS** as an urgently needed paradigm shift – from complacent preservationist views to a proactive security-orientated approach – and provide a structured appendix detailing the index methodology (indicator list, normalisation, scoring formula, weights, and thresholds) to foster adoption and refinement by researchers and policymakers.

**Keywords:** Arabic- International language Databases- UNESCO- Ethnologue’s EGIDS Framework- Danger of Extinction- linguistic Erosion- National Security- Social Cohesion- Early Warning

## INTRODUCTION

The contemporary status of the Arabic language presents a stark paradox. International language databases and indices routinely classify Arabic among the world’s most secure and robust languages. UNESCO celebrates Arabic as a pillar of cultural diversity, used daily by hundreds of millions, while Ethnologue’s EGIDS framework accords Standard Arabic the highest vitality rating (EGIDS 0, “International”) reserved for global lingua francas. By traditional metrics – sheer number of speakers, official recognition, liturgical importance – Arabic appears in little danger of extinction. Yet such assessments belie a “tumultuous sea of threats” beneath the surface. This paper advances a countervailing thesis: Arabic is undergoing gradual but accelerating attrition, a decline in its functional domains and prestige that, while not immediately terminal, poses a strategic threat to the fabric of Arab societies.

This paper argues that this linguistic erosion is not a benign cultural shift but a national-security-relevant risk factor. Rather than positing a simple, immediate causal chain from Arabic decline to violent conflict, the paper conceptualises language attrition as one of several structural dynamics that can weaken social cohesion, deepen class and regional cleavages, and complicate state-society communication over the medium to long term. Comparative work on Pakistan's Urdu-Bengali tensions<sup>1</sup>, Sri Lanka's Sinhala-only reforms, and the politicisation of Serbo-Croatian varieties in late Yugoslavia suggests that when language hierarchies come to overlap with entrenched grievances of identity and inequality, they can become powerful accelerants of fragmentation. While these cases are not directly analogous to the Arab world, they show that shifts in the functional status of a major language can interact with other drivers of instability in ways that are strategically consequential. Because Arabic remains widely transmitted across the region, we do not claim that it has already precipitated instability; rather, we argue that current patterns of domain loss and stratified multilingualism warrant being treated as early warning signs of potential security-relevant tensions.

At present, however, the prevailing tools to assess language health grossly underestimate the severity of Arabic's attrition. Conventional models like UNESCO's World Atlas of Languages (WAL) and the Expanded Graded Intergenerational Disruption Scale (EGIDS) focus on static metrics – e.g., absolute speaker numbers, presence in education, intergenerational transmission – that paint Arabic as “safe” because children still learn it as a mother tongue across the Arab world. These models are lagging indicators of language decay. By the time a language drops a tier on EGIDS (e.g., from “safe” to “threatened”), it has already lost irretrievable ground. In Arabic's case, the alarm bells of traditional indices won't ring until very late – likely only after younger generations have largely switched to another language at home. We contend that waiting for such late-stage symptoms (like failing parent-child transmission) is untenable. The true threat to Arabic is not sudden death but slow attrition: an accumulating loss of domains (science, commerce, education, media) and a waning of perceived utility, occurring here and now even as raw speaker numbers remain high.

This paper develops an alternative approach commensurate with the nature of the threat. We introduce SPH-LENS – the Socio-Political-Historical Language Early-warning & National-security System – as a framework to monitor and preempt Arabic language decline. SPH-LENS builds on a theoretical foundation that integrates sociolinguistic capital (the value and utility of a language in socio-economic life) and linguistic hegemony (the soft power by which a dominant language displaces others) within a historical and geopolitical context. By structuring the problem along three dimensions (S, P, H), we capture the multifaceted drivers of attrition: from youth attitudes and digital usage (socioeconomic) to state policies and cultural narratives (political) to colonial legacies and spatial dynamics of the Arab world (historical). Crucially, SPH-LENS is indicator-driven and forward-looking. It operationalises each dimension with concrete metrics drawn from existing data sources – for example, tracking the proportion of scientific publications or web content in Arabic, the prevalence of English in higher education and business, the strength of pro-Arabic language policies, and so on. By aggregating these metrics into a composite risk index (with transparent normalisation and weighting), the framework functions as an early-warning system: a low score or sharp downward trend would “red-flag” a country's language vitality before the point of no return.

The intended contribution is twofold. First, the paper fills a critical gap in language vitality research by proposing a model tailored to detect dynamic decline in a major lingua franca – a scenario for which existing endangerment indices (designed mainly for minority languages) are ill-suited. Second, we recast language preservation as strategic policy imperative. Rather than viewing Arabic's fate as a cultural concern for linguists or educators alone, we frame it as a core component of national security planning. SPH-LENS is designed for practical usability by governments and international organizations: it relies on publicly available data (from UNESCO, World Bank, etc.) and offers a replicable methodology that can be updated regularly to monitor trends. By demonstrating SPH-LENS with real-world cases – such as the domain shifts in Tunisia's and Morocco's education systems<sup>2</sup>, or the dominance of English in UAE's private sector<sup>3</sup> – we aim to show how this framework can illuminate risks that would otherwise slip under the radar.

In what follows, we first review the theoretical background that underpins SPH-LENS, including the concepts of sociolinguistic capital, hegemony, and historical contraction of language domains. We then detail the limitations of conventional vitality models in the Arabic context, to underscore why a new approach is necessary. The core of the paper presents the SPH-LENS framework: its three dimensions, selected indicators,

and how these combine to yield a composite early-warning index (**with full methodological details provided in the Appendix**). We compare the insights from SPH-LENS against those from UNESCO and EGIDS classifications, highlighting how our framework captures early signs of threat that static models miss. Finally, we discuss policy implications: how Arab states can use SPH-LENS to inform strategic interventions – from education reforms to media and technology policies – that reinforce Arabic’s role, thereby safeguarding social cohesion and security. The conclusion calls for a paradigm shift in how we value and defend the Arabic language, advocating the adoption of tools like SPH-LENS as part of a broader “linguistic security” agenda for the Arab world.

## Theoretical Background

Language is more than a medium of communication; it is a vessel of identity, social memory, and political unity. Throughout history, states have recognized that a common language undergirds national cohesion – and conversely, that linguistic fragmentation can destabilize nations. The Arab world’s own historical narrative testifies to this. Classical Arabic, through the Qur'an and a transnational scholarly tradition, bound together a vast civilization. In modern times, Arabic’s role in pan-Arab identity has been explicitly acknowledged by Arab thinkers and leaders. Its potential erosion therefore raises alarms akin to losing a critical infrastructure: when a shared language fades, societal fractures widen, and external actors may more easily exert influence via the ascendant languages.

Comparative experience underscores both the potential and the limits of these linkages between language regimes and political stability. In Pakistan, for example, the post-independence elevation of Urdu over Bengali contributed to grievances in East Pakistan and is widely cited as one factor in the eventual secession of Bangladesh. In Sri Lanka, Sinhala-only language reforms exacerbated Tamil marginalisation and fed into decades of civil war<sup>4</sup>, while in the late Yugoslav federation, the rebranding of Serbo-Croatian into separate national standards both reflected and deepened the pull of competing national projects<sup>5</sup>. None of these crises can be reduced to language alone – economic inequality, territorial disputes, and external intervention all played decisive roles – but they illustrate how the symbolic and practical ordering of languages can interact with other drivers of contention to either mitigate or magnify fragmentation.

Other multilingual polities demonstrate that linguistic diversity can also be sustainably institutionalised and even turned into an asset. Switzerland, Canada, and several postcolonial states have developed arrangements in which multiple languages coexist within a single political community and provide bridges to different economic and diplomatic spheres<sup>67</sup>. The problem, then, is not multilingualism per se but stratified multilingualism: situations in which one language is associated with high-status education, mobility, and power while another is relegated to low-prestige or purely symbolic functions. SPH-LENS is designed to identify precisely these stratified configurations in Arab states. It does not assume that the growing presence of English or French is inherently destabilising but asks when and how the cumulative marginalisation of Arabic from high-value domains might increase the vulnerability of already fragile social contracts.

It is with this understanding that we elevate the discussion beyond preserving culture for culture’s sake. Preserving Arabic is about preserving national integrity and resilience. When youth in Tunisia or the UAE begin to regard Arabic as “dated” or irrelevant, preferring English as the language of science and progress, it heralds a strategic inflection point. The early warning signs of language endangerment, as linguists have noted, include exactly this kind of functional replacement: when English or French displaces Arabic in literature, education, or commerce, and when younger generations feel their mother tongue is unsuitable for modern life. These signs are apparent today: e.g. in the UAE, observers note a “continual erosion of the Arabic language and its substitution with English” in daily business and even at home. Such trends, if unchecked, forecast a scenario where Arabic could eventually retreat to informal or religious domains only, no longer the language of power in its own homelands. That outcome would weaken the sociopolitical glue of the Arab states and could fuel internal disparities (an Arabic-speaking underclass vs. an English-educated elite) that adversaries might exploit.

## Limitations of Traditional Vitality Models

Why have mainstream language vitality assessments failed to register this looming threat? The crux of the issue lies in what is measured. UNESCO's World Atlas of Languages (WAL) and Ethnologue's EGIDS are invaluable for cataloguing endangered minority languages, but their criteria are too narrow for a diglossic, rapidly changing context like Arabic. These models emphasise intergenerational transmission: a language is "safe" as long as children learn it as a first language and use it in daily life (EGIDS Level 6a or above). By that yardstick, Arabic – still spoken in homes from the Atlantic to the Gulf – scores very high. But the models have given short shrift to functional loss: the erosion of a language's use in high-prestige domains even as it remains spoken colloquially.

In Arabic-speaking societies, children may grow up fluent in the dialect of Arabic for everyday matters (ensuring a high EGIDS rating), yet those same children might attend schools where science and maths are taught in French or English, consume primarily English-language media, and later work in jobs where reports are written in English. The traditional indices wouldn't mark Arabic as "threatened" until perhaps generations later, when those children might, in the worst case, not teach Arabic to their own kids. By then, the "foundation would have decayed" – the language's utility and desirability would have been so diminished that its marginalisation is almost irreversible.

Consider the UNESCO Atlas of Endangered Languages (the precursor to WAL). It does not list Arabic at all – understandably, since Arabic is not on the brink of extinction like an indigenous tongue with a few hundred speakers.<sup>8</sup> Yet Arabic could be undergoing a slow status death: thriving in raw speaker numbers but dying in the higher registers of usage, what some scholars call "language contraction". Existing frameworks lack a means to quantify this. The EGIDS has a category for languages used only in certain domains (e.g. an academic lingua franca), but as a hierarchical scale it forces Arabic into the top tier by virtue of its official status and wide use. It cannot express that within that top tier; Arabic's domain strength might be weakening year by year.

Moreover, these models largely ignore sociopolitical drivers. They are rooted in the sociolinguistics of small communities, where the key concern is whether parents transmit the mother tongue. For Arabic, the more relevant questions are: Why might parents (or students) choose not to fully transmit or develop Arabic? What pressures or incentives lead them to favour another language? The answers lie in socioeconomic calculations and power dynamics that UNESCO/EGIDS barely incorporate. For instance, the EGIDS description for a "National" language (Level 1) assumes it is used in education, work, mass media, etc. – which is true for Arabic on paper. But if, in reality, the elite jobs require English and the best university programs are delivered in English, parents will indeed begin to see Arabic-medium schooling as a disadvantage, no matter what the official policy states. By focusing on whether Arabic is still spoken at home, EGIDS misses the gradual motivational shift that precedes a break in transmission. In many Arab cities today, educated families still speak Arabic at home (keeping EGIDS at 1), yet they are simultaneously investing heavily in English education for their children and even peppering their own Arabic with English technical terms. The "tipping point" for language shift is brewing not at home but in the workplace and school.

A telling statistic was highlighted in the Arab Human Development Report (2003): over three decades, only about 10,000 books were translated into Arabic, roughly equivalent to what Spain translates in one year<sup>9</sup>. This was cited as evidence of a knowledge deficit and a "crisis in contemporary Arab culture." From a language vitality perspective, it's an indicator that Arabic is not the language of access to new knowledge for Arab populations – foreign languages are.<sup>10</sup> Yet none of the mainstream vitality indices would register this knowledge import gap. Likewise, internet presence is a critical modern vitality sign. Arabic is among the top languages by number of users online, but Arabic content constitutes only about 0.6% of indexed websites. An Arabic speaker can use the internet, but much of what they read may be in English or other languages. Again, traditional models have no mechanism to factor in the digital realm (UNESCO is only now beginning to discuss "cyberspace" in language status terms).<sup>11</sup>

## Integrating Sociolinguistic Capital and Hegemony (SPH Foundations)

The SPH framework stands on three theoretical pillars corresponding to its initials:

- **Socioeconomic (Sociolinguistic Capital):** The paper draws on Pierre Bourdieu's notion of linguistic capital – the idea that language functions as a form of cultural capital that can confer access to social and economic rewards. In any given “linguistic marketplace”, different languages have different market values. Individuals (and institutions) will rationally invest in the language that offers the best return (in jobs, education, and status). Historically, Arabic carried enormous linguistic capital in the Arab region – it was the language of religion, literature, law, and identity, with knowledge of Arabic yielding respect and opportunities. However, the past century (especially the past few decades of globalisation) has seen a dramatic revaluation: English (and regionally French) have appreciated in value, while Arabic's relative value has depreciated in key sectors. Our earlier study illustrated this with a corporate example: among top Arab companies of the older generation, 70% had Arabic names, whereas among new tech startups, only 6% did. Fluency in English is increasingly a prerequisite for high-paying jobs (98% of emerging Arab fintech companies required English for senior roles). These figures indicate that English now carries higher linguistic capital in the Arab economic sphere than Arabic. Bourdieu's framework predicts what we indeed observe: a feedback loop where people invest more in English (through education, code-switching, etc.) to reap its rewards, further eroding Arabic's capital and usage.

In SPH-LENS, the Socioeconomic dimension (S) aims to capture this dynamic. The paper includes indicators that reflect the market value of Arabic vs. other languages: for instance, education and research outputs in Arabic (Are universities producing theses and papers in Arabic, or only in English? How many scientific publications from Arab countries are published in Arabic journals – as can be gauged via databases like OpenAlex?) digital presence (What share of web content or popular media in a country is in Arabic, per W3Techs or Wikimedia stats?), and human capital metrics (What are English proficiency levels, per EF EPI, and is there an inverse correlation with Arabic usage? How many students study abroad in Anglophone/French institutions, potentially shifting their linguistic habits. Each of these speaks to Arabic's perceived utility. For example, the EF English Proficiency Index rankings show several Arab countries with improving English skills, like Morocco and Tunisia achieving “moderate” proficiency. On one hand that reflects positive education outcomes, but from an Arabic perspective it may indicate that large segments of society are orienting toward English for professional advancement.

In Morocco, 65% of young people consider English important to learn, slightly more than those who value Arabic (62%) – a striking finding that English is seen as even more vital than their own language for the future.<sup>12</sup> These statistics measure the sociolinguistic capital gap: when English is prized more than Arabic by Arab youth, it quantifiably signals attrition risk. SPH-LENS uses such data to score the Socioeconomic dimension, effectively asking: To what extent is Arabic maintaining vs. losing its role as a language of opportunity and everyday utility?

- **Political (Hegemony and Policy):** The second pillar comes from Antonio Gramsci's concept of cultural hegemony, which we apply to language. Gramsci showed how a dominant group's worldview can become “common sense” for all, without force, by infiltrating culture and ideology. Linguistic hegemony describes how one language can achieve dominance by being seen as the natural vehicle of modernity, power, and progress. In the Arab world, colonial history and globalization have established English (and to a lesser extent French) as hegemonic in many domains. Crucially, this dominance is self-reinforcing and largely structural. It's not that Arab parents or students have personally turned against Arabic; rather, they operate within a system where English is rewarded. Government policies often inadvertently abet this. For instance, many Arab countries maintain dual tracks in education – elite private schools and universities where the medium of instruction is English (or formerly French), versus mass public schools in Arabic. The former track is associated with prestige and upward mobility, feeding a class divide where the elite perpetuate English dominance.

Even state institutions sometimes prioritize international languages for outward-facing functions: science research is published in English to reach journals; tourism and business materials in Gulf states are often

English; some legal contracts might be in English especially in business zones. These choices cumulatively send a message that Arabic is provincial, English is global – a classic hegemonic narrative. Over time, the dominated group (in this case, Arabic speakers) may internalize a sense of their language's inferiority. It becomes “common sense” that serious knowledge or high-tech innovation happens in English, not in Arabic. This devaluation is exactly what has been happening: in survey data, large majorities of Arab youth refer to English as “the language of science, of business, of the future”. That encapsulates linguistic hegemony at work.

The Political dimension (**P**) of SPH-LENS encompasses both the overt policy environment and the subtler aspect of hegemonic pressure. To quantify it, we include indicators such as the constitutional and legal status of languages (from the Constitute Project or national laws – e.g. does the constitution mandate Arabic for all official business? Are there language laws about signage, media quotas, etc.? Such measures show formal commitment to Arabic), language of instruction policies (e.g. whether STEM subjects in secondary/tertiary education are taught in Arabic or a foreign language, which could be gleaned from UNESCO or national stats), media and cultural production in Arabic (perhaps using UNESCO's statistics on domestic film/book production or the share of foreign vs local language media content, and freedom and governance indices (like Freedom House scores or the World Bank's GovTech Maturity Index). The latter may seem indirect, but the hypothesis is that governance factors influence language outcomes. For example, a high GovTech Maturity Index might indicate advanced e-government – we can then ask, is this e-government accessible in Arabic or primarily in English? (A country could score high by digitizing services but using English interfaces, which would highlight a gap.) Conversely, a country with strong political will to preserve Arabic might implement comprehensive Arabic digital platforms regardless of GTMI ranking<sup>13</sup>. Freedom House's metrics on media freedom might correlate with whether local language content thrives or whether foreign media dominate the space. Also, an authoritarian regime might enforce Arabic usage (as a nationalist tool), whereas a very open market might allow English to flood in unregulated. Thus, the political dimension's indicators combine to assess the structural power of Arabic in society: is Arabic institutionally promoted and privileged, or is it effectively sidelined by the policies and power relations that favor another language? For instance, Morocco's recent law switching the instruction of science back to French (after years of Arabization) is a policy choice with huge implications. SPH-LENS would mark that as an unfavorable development in the Political dimension, reflecting a state-level decision that diminishes Arabic's domain. On the other hand, the UAE's high-profile initiatives like the “State of the Arabic Language” report and plans to boost Arabic digital content might score positively as efforts to counter hegemony with conscious policy.

- **Historical (Legacy and Contraction):** The third pillar recognizes that Arabic's situation today cannot be divorced from historical context – both the historical trajectory of Arabic itself and the particular histories of each Arab country vis-à-vis language. We refer to “historical contraction” to denote the long-view trend of Arabic's changing expanse. After the explosive spread of Arabic with the early Islamic conquests (7<sup>th</sup>–8<sup>th</sup> centuries) and its golden age as the lingua franca of a vast civilization, Arabic eventually ceded ground in some areas (e.g. Al-Andalus with the Reconquista, or the rise of other regional lingua francas under colonialism). In the modern era, roughly since the 19<sup>th</sup> century, Arabic has faced pressure from European languages due to colonization and globalization. The peripheral regions of the Arab world – North Africa (Maghreb) and the Mashreq periphery – were under European rule longer and more intensely, embedding French or English deeply. Meanwhile, the core Arab lands (the Arabian Peninsula and parts of the Levant) retained Arabic more exclusively in public life until the oil age/globalization wave. This yields a hypothesis, borne out by data in our prior work: countries geographically or historically more distant from the “Arabic heartland” show faster attrition. For example, in Somalia and Djibouti (on the fringes of the Arab League), Arabic is culturally less entrenched and faces competition from Somali, French, etc.; in North African states like Tunisia or Morocco, a century of French dominance in education left a lasting imprint such that French (now increasingly English) is still seen as the key to success. In contrast, Saudi Arabia or Yemen have no equivalent colonial legacy; the struggle there is more with the recent influx of English via globalization.

The Historical dimension (**H**) thus includes factors capturing legacy influences and structural inertia. One component is colonial legacy metrics: whether a country was a French or British colony/protectorate and for

how long, which often correlates with the continued use of French or English. Algeria and Morocco, for instance, spent over 40 years post-independence trying to “Arabize” education that had been in French; the mixed success of those efforts is a legacy effect measurable in today’s bilingual proficiency rates. We can use data like the percentage of people proficient in the ex-colonial language as a proxy (e.g. Francophone vs. Anglophone penetration). Another factor is the linguistic demographics and minority languages historically present: countries with significant non-Arabic native languages (like Sudan with Nubian, or Morocco with Berber/Amazigh which is now an official language alongside Arabic) have a different dynamic than fully Arabic-speaking populations. The presence of a strong second indigenous language can either dilute Arabic’s dominance or, conversely, if that second language is also under pressure, Arabic might unify (each case requires careful reading of indicators like linguistic diversity index, available via Ethnologue or UNESCO). We also consider historical educational data: literacy rates in Arabic over time, the timeline of when mass education in Arabic started (some Gulf states only established modern Arabic curricula mid-20th century; earlier elites studied in English abroad). These historical patterns influence how resilient Arabic is in each society’s psyche. Additionally, this dimension accounts for geospatial factors: for example, distance from the traditional centers of Arabic literary production (Cairo, Damascus, Baghdad historically). Our previous model quantified a “peripherality index” – the further a country from the Arab core, the more at risk, partly due to closer interface with other linguistic spheres. While such a metric isn’t a standard global index, we use it conceptually: e.g. Mauritania or the Comoros, being on the extreme edge, might get a flag in our analysis as places where Arabic’s hold could be tenuous. Concretely, the Historical dimension in SPH-LENS might incorporate data like: years since independence from non-Arab rule, official secondary languages recognized (e.g. if French is co-official or formally used, as in Djibouti or Somali), and continuity of Arabic in government/education (did a country break continuity by adopting another language for a significant period? Morocco’s flip-flop between Arabic and French in science instruction is a case of historical back-and-forth that affects generational competence in Arabic for scientific discourse).

Together, these three dimensions (**S, P, H**) form a holistic lens to examine Arabic language vitality. Each dimension is not isolated; they interlink – hegemony (**P**) erodes linguistic capital (**S**), and both play out differently given a nation’s history (**H**). By structuring our framework this way, we acknowledge that no single metric can signal attrition risk, but a composite picture can. For instance, Tunisia might score moderately on Sociolinguistic Capital (because many still speak Arabic natively and Tunisia has high Arabic literacy) but poorer on Political (due to the strong role of French in higher domains) and its Historical legacy of French education. Saudi Arabia might be the opposite: strong historical/core position and strong official status for Arabic (Political) but facing socio-economic pressures (**S**) as it internationalizes and brings in English curriculum in universities. SPH-LENS is designed to pick up such nuances, translating them into a comparative index of “Arabic attrition risk”.

## The SPH-LENS Framework: Indicators and Methodology

To transform the SPH theoretical model into a practical monitoring tool, we identify a set of quantifiable indicators for each of the three dimensions (Socioeconomic, Political, Historical). These indicators are drawn from existing global and regional datasets, ensuring that SPH-LENS can be populated with regularly updated, publicly available information. In this section, we outline the core indicators under each dimension, justify their significance, and describe how they collectively inform an early-warning index of Arabic language attrition risk.

### Socioeconomic Dimension (S): Measurable Signs of Linguistic Capital Erosion

#### Education Medium and Attainment in Arabic

One of the clearest indicators of Arabic’s practical vitality is its role in education. For each country, we examine the language of instruction in secondary and tertiary education. Data from UNESCO Institute for Statistics (UIS) and national education ministries often detail whether STEM subjects in secondary school are taught in Arabic or a second language, and what the primary medium in universities is. For example, Algeria and Syria use Arabic for most university programs (with some exceptions), whereas Qatar and the UAE use English in many technical fields. A decreasing presence of Arabic as the medium of instruction, especially in

scientific and technical disciplines, is a red flag. Additionally, student achievement tests (like TIMSS or PISA scores) segmented by test language can be revealing: countries where students take these tests in English or French (rather than Arabic) indicate a de facto shift (Morocco, for instance, began allowing French on some standardized tests after reintroducing French in science classes).

## Scientific and Scholarly Output

The production of knowledge in Arabic vs. other languages is a strong marker of linguistic capital. Using bibliometric databases like OpenAlex (which catalogs academic publications globally) or Scopus, we can estimate the proportion of research papers by Arab authors published in Arabic. OpenAlex allows filtering by country and language of publication. If, for example, less than 5% of research articles originating in Egypt are written in Arabic, with the rest in English, it signifies that even at the highest levels of intellect, Arabic is sidelined. Indeed, a *Nature* analysis in 2023 noted that Arabic ranks low among languages of translation and scientific publication, with far more works translated from English into Arabic than original works produced in Arabic.<sup>14</sup> Another proxy is the volume of Arabic academic journals and their citations. The UNESCO Arabic Science Report (if available) or data from the Arab League's Educational, Cultural and Scientific Organization (ALECSO) might enumerate scientific publications in Arabic. A downward trend or small share in this area indicates domain contraction.

## Digital Presence and Media Use

In today's world, digital content is crucial. We incorporate W3Techs statistics on content language to quantify Arabic's share on the web. As of late 2025, only about 0.5% of websites have Arabic as the content language, a minuscule fraction considering Arabic speakers are ~5% of the world population. While some popular platforms (Facebook, YouTube) are multilingual, the low figure implies Arabic-speaking internet users either consume English content or are underserved. Tracking this over time is instructive: W3Techs provides historical trends, and Arabic's percentage has fluctuated (with some increase last decade due to more Arabic social media content, but still under 1%).<sup>15</sup> Another digital indicator is Wikipedia activity. The number of articles and active editors on Arabic Wikipedia versus those on English Wikipedia from Arab countries can be gauged via Wikimedia stats. If educated Arabs prefer contributing in English (or consuming English Wikipedia), it reflects language preference. Furthermore, we look at social media usage patterns: surveys or studies sometimes reveal what languages people use on Twitter or Instagram in the Arab world. A strong preference for English in professional or public social media discourse, as has been observed in the Gulf, marks a shift in prestige.

## English Proficiency and Use

The EF English Proficiency Index (EF EPI) is a standardised measure that ranks countries by the English skills of adults. Most Arab countries rank low to moderate, but importantly the trend is upward in many places, meaning English competence is rising among younger generations. For instance, EF EPI 2021 ranked Tunisia and Morocco highest in North Africa (both "low proficiency" but nearing moderate), and the UAE was in the "moderate" band. We interpret a high EF EPI score as both cause and effect of Arabic attrition: cause, because high English proficiency often comes from an education system that prioritises English (perhaps at Arabic's expense); effect, because once people know English well, they may use it preferentially for technical subjects or online activities, reinforcing English dominance. However, EF EPI alone could be misleading – some countries with low English (e.g., Iraq, Yemen) are not necessarily safe for Arabic, since other factors (like poor education quality or dominance of another second language) could be at play. So, EF EPI is used in combination with other indicators. We also note English usage in business (e.g., the percentage of job postings requiring English, which can be gleaned from labor market surveys or even LinkedIn data). In the earlier example, 98% of new Arab fintech companies required English for management. That kind of data, when available, is a direct metric of how indispensable English has become vis-à-vis Arabic in the economy.

---

## Intellectual Property and Innovation

As a proxy for the language of innovation, we use WIPO statistics on patent applications and trademarks in the Arab region. Specifically, whether applicants file in Arabic or another language. Many Arab countries allow patent filing in English or French (the GCC Patent Office, for example, has historically published patents in English with Arabic bibliographic data). If domestic innovators themselves choose English for filing patents or registering trademarks, it underscores that cutting-edge work isn't being encapsulated in Arabic. Quantitatively, we might look at the number of international patent applications (PCT) originating from Arab states and check how many of those were submitted in Arabic via national offices. Anecdotally, this number is very low, since even Arabic-speaking patent attorneys often draft in English to ease international processing. While not a commonly cited indicator in linguistics, we include it to reflect high-level usage of language in the knowledge economy.

## Economic and Demographic Signals

Broader socioeconomic data, such as urbanisation rates and migration patterns, also feed into language usage. Urban elites tend to adopt global languages faster; large-scale emigration (brain drain) to Western countries can increase the prestige of English/French back home through diaspora influence. The World Bank's Global Talent Mobility indicators (or similar metrics on student migration and expat communities) give context. For example, hundreds of thousands of Arab students study abroad in English or French each year; when they return, they may be less likely to work in an Arabic-medium environment. The number of international schools and foreign universities in-country (e.g., branch campuses in Qatar's Education City, English-medium universities in Egypt or the UAE) is another sign – we track these as a count per country, indicating parallel education structures that compete with Arabic institutions.

Collectively, the Socioeconomic indicators provide a data-driven picture of Arabic's practical vitality: whether the average young person in an Arab country can and does live their advanced educational, professional, and digital life in Arabic, or whether they must switch to a foreign language to access opportunities. A strong Arabic vitality would mean high metrics for Arabic-medium education, Arabic content, and low reliance on English for success; an at-risk scenario is the reverse.

## Political Dimension (P): Policy, Power, and Hegemonic Influence

### Official Language Status and Constitutional Provisions

All Arab League states declare Arabic as an official language, but some constitutions go further than others in specifying its usage. Using data from the Constitute Project, we examine each constitution for clauses on the Arabic language – e.g., is Arabic mandated as the language of government proceedings, education, and national media? Some constitutions, like Algeria's, also recognise Tamazight (Berber) as official, and others note the protection of French/English usage in certain contexts. A simple indicator is a binary “Arabic is the sole official language (yes/no)”. Most are “yes”, but the nuance lies in any official secondary languages (Somalia recognises Somali and Arabic; Iraq recognises Kurdish regionally, etc.). A more telling metric: the constitutional rank of Arabic – primary official, co-official, or just national language without enforcement. We assign a higher score to countries giving Arabic exclusive or primary official status in law, under the assumption that it reflects stronger institutional support.

### Language Policy and Planning Efforts

This includes any national language councils, laws or programmes aimed at promoting Arabic. For instance, does the government have regulations on Arabic signage (like Qatar's law that Arabic must be prominent on business signs), quotas for Arabic content on TV/radio, requirements for civil servants to be proficient in Arabic, etc.? Presence of such policies indicates proactive maintenance. Conversely, absence or lax enforcement suggests a laissez-faire approach that might allow rapid incursion of other languages. We can derive a qualitative score or count of language policy measures per country (sources include UNESCO reports and national legal databases). For example, Saudi Arabia in 2020 launched a new Arabic Language Academy

and initiatives to ensure Arabic use in government documents – these would count as positive policy actions. Meanwhile, Lebanon has almost no enforcement of Arabic in the private sector (French and English freely dominate advertising, etc.), which would be a weaker policy environment for Arabic.

### **Freedom of Language Use vs. Cultural Protection**

This indicator is somewhat complex: using Freedom House indices (on civil liberties and press freedom) to gauge whether a country's openness allows foreign media and schools to proliferate, or if it takes a controlled approach, possibly protecting Arabic. Interestingly, some of the Gulf states with lower Freedom House scores implement strong Arabisation in public signage and media (as part of nationalist policy), whereas freer countries might not regulate language use at all. We interpret extremes carefully: a very low freedom score might correlate with high Arabic usage in official domains (because the regime mandates it), but that could coexist with English dominance in unofficial domains, especially if the elite circumvent restrictions. A high freedom score might correlate with high penetration of Western media and education, thereby more English. Neither extreme is purely "good" or "bad" for language vitality. Instead, we look for balance: e.g., Tunisia post-2011 became freer, and simultaneously there's been vigorous debate and activism about replacing French with English in schools – indicating both the influence of open discourse and the risk of simply swapping one hegemonic language (French) for another (English) rather than bolstering Arabic. We include Freedom House's "Freedom on the Net" as well: a country with an open internet environment will have unfiltered access to English content (which is fine for freedom, but from a language perspective, means Arabic content competes with a deluge of English). In sum, we don't judge a free society negatively, but we use these metrics to contextualise other data (for instance, understanding that a fully open media market in Jordan means Hollywood and Egyptian Arabic content flood in, overshadowing local Arabic dialect content, etc.).

### **International Hegemonic Pressures**

Indicators here might be less formal. We consider foreign cultural institutes and schools (number of French lycées, British/American schools, Goethe/China cultural centres, etc. in the country). A higher presence can indicate stronger soft power from non-Arabic cultures. For instance, Morocco and Lebanon have many French and American schools, a legacy of foreign influence, whereas Yemen has very few. Similarly, membership in international anglophone/francophone blocs (like the Francophonie for some North African states as observers) can be a factor. We also note if the country has adopted international exams in a foreign language (many Gulf high schools use SAT/AP or IB in English). All these hints at how much the local elite align with external linguistic standards versus local ones.

### **Government Tech and Services Language**

Using the World Bank GovTech Maturity Index (GTMI) in a novel way, we pair it with an analysis of the languages offered in e-government services. GTMI classifies countries into tiers of digital government advancement. The UAE, Saudi Arabia, and Bahrain rank highly in GovTech, meaning citizens can do many services online. We check: are those services offered in Arabic by default? (In the GCC, yes, typically both Arabic and English interfaces are provided.) Countries with lower GovTech might not have much online content, ironically preserving in-person Arabic use but also possibly meaning people rely on foreign platforms for info. If a high-GTMI country were to neglect Arabic in its digital services, that would be a glaring issue – so far, that's not generally the case in Arab countries, but it's worth monitoring. The reason to include GTMI is that as governments digitise, ensuring Arabic digital inclusion becomes key. If, say, an Arab government rolled out an e-service only in English (perhaps using off-the-shelf software), that would directly diminish Arabic's functional domain. Our framework would flag that via this indicator.

### **Media and Press Language**

The paper measures how much of the major media (TV channels, newspapers, news websites) is in Arabic vs. foreign languages. Many Arab countries have prominent English or French newspapers (e.g., The Daily Star in Lebanon (English), Jordan Times, Arab News in Saudi, Egypt Independent English edition, etc.) and TV channels (Dubai has an English TV; many countries have French channels for their Francophone minority). A

high ratio of foreign-language media consumers suggests an elite or expat community not using Arabic media. For example, in the UAE, newspapers in English like Gulf News or Khaleej Times have significant circulation among certain demographics, and an English-language state news channel exists. We use data from audience surveys or media reports about readership/viewership. UNESCO's media statistics might not break by language, but national statistics sometimes do. We also consider internet news consumption – e.g., trends like Arab youth increasingly following international English news on social media rather than Arabic news sources. If a society's information sphere tilts away from Arabic, that's a strategic vulnerability (public opinion and discourse happening outside the Arabic language).

## **Historical Dimension (H): Legacy Factors and Long-Term Trends**

### **Colonial Legacy Index**

The paper assigns a score based on the length and type of colonial rule that impacted language. A simple scale: 0 = never colonised by a Western power that imposed a new language (e.g., Saudi Arabia, which had only a brief indirect British influence, or perhaps most of the Gulf states before oil); 1 = under colonial mandate, but Arabic remained central (e.g., Iraq under short-term British rule); 2 = heavy colonial linguistic influence (Maghreb under French, where French became the language of administration and education for decades). Algeria, Morocco, and Tunisia would score high on colonial legacy impact, as French became deeply ingrained (so their "Arabic Historical resilience" is lower).

This index is qualitative but supported by metrics like year of independence minus year of colonisation = number of years a foreign language was official. Algeria (1830–1962 French rule, ~132 years) vs. Saudi (never formally colonised) exemplify the extremes. The index correlates with present bilingualism rates – e.g., Morocco's population today still widely speaks French, reflecting that legacy.

### **Historical Arabization Policies**

This is almost the converse of the above: after independence, how vigorously did the state restore or promote Arabic? Countries like Morocco and Algeria undertook "Arabisation" of administration and schools in the 1960s-1980s, albeit with mixed success (Morocco partially reversed it in 2019 for sciences). Syria and Egypt, though not colonised in the same way, also modernised Arabic terminology and curriculum in the mid-20th century (Egypt had Arabisation in some university programmes post-1952). We compile whether a country had a formal Arabisation campaign and whether it was sustained. If yes and successful (e.g., Algeria today conducts most schooling in Arabic), that gives some inertia in favour of Arabic – though ironically, Algeria's university system reverts to French for scientific fields, undermining it. Conversely, a country that did not need Arabisation (because it was always Arabic-centric) might be assumed strong historically, but one must watch if they now are adopting foreign languages (the Gulf states fall here: historically fully Arabic in identity, but currently very exposed to English).

### **Linguistic Core vs. Periphery**

The paper utilises a geographical-cultural distance factor. The "core" could be thought of as the Arabian Peninsula and perhaps Cairo/Damascus as traditional intellectual centers. The farther away (Northwest Africa, East Africa) or more culturally distinct (with non-Arab ethnolinguistic groups), the more likely attrition. For instance, Somalia and Djibouti: Arabic is official, but Somali and Afar are native tongues; Arabic fluency is not universal. These are edge cases in the Arab League where Arabic's vitality relies on pan-Arab sentiment rather than mother tongue usage – so they rank as high risk historically. A measurable angle: the percentage of population whose first language is Arabic. In most Arab states it's ~99%, but in ones like Somalia, Comoros, Djibouti, that percentage is lower (Somalia maybe ~35% speak Arabic as second language mainly; Comoros has Comorian as primary).

Similarly, Sudan historically had English during Anglo-Egyptian rule and many local languages; South Sudan split off, partly due to linguistic-cultural divides. In our index, Sudan would carry some historical vulnerability

because Arabic was not the sole heritage language countrywide. The paper gathers such data from Ethnologue and censuses: if less than, say, 80% of the population speaks Arabic natively, we mark that.

## Past Trends in Arabic Usage

If historical data exists (from older surveys or qualitative accounts) about earlier generations' use of foreign languages, we can establish a trendline. For example, Tunisia's older elite (1960s) spoke French dominantly in many domains; after independence, Arabic made gains, but now English is rising – a nonlinear trajectory. We incorporate known points: printing and publishing in Arabic across decades, number of Arabic newspapers in 1950 vs 2000, etc., to see if there was a golden age and then decline or vice versa. Countries like Egypt have seen a long-term increase in Arabic print media until recently (when digital disruption came, and English digital content competes).

## Cultural Continuity (Religious and Literary)

One historical strength of Arabic is its role in religion (Islam). Societies with strong Islamic education may maintain classical Arabic knowledge irrespective of modern trends. We consider data like Qur'an schooling rates, presence of traditional madrasa system, etc. This might seem outside modernization, but it's relevant: for instance, Mauritania has a very strong tradition of Arabic-Islamic scholarship (high per capita Hafiz, etc.), which historically preserved Arabic even under French colonization. That gives Mauritania some resilience (though economically it's weak and French still is important). On the other hand, countries where secular education dominated and Arabic classical learning declined might have fewer deep roots (e.g. in some Marxist period South Yemen minimized religious schooling, possibly weakening classical Arabic familiarity among that generation). The paper use proxies like number of religious Arabic schools or proportion of population that has attended them.

## Previous Signs of Shift

Finally, any recorded incidents of language shift or public controversy historically are telltales. In Morocco, the "Arabisation vs. French" debate is decades old; in the UAE, concerns about Arabic decline have been voiced at least since the 2000s. If a country has already identified the issue and perhaps attempted remedies (like Qatar's 2012 policy to reintroduce Arabic in university instruction after a period of English-only at Qatar University), those events are catalogued. They show either resilience (a correction was made) or persistent struggle.

By quantifying these historical and legacy factors, SPH-LENS doesn't treat all Arab states as starting equal. It acknowledges, for example, that Maghreb states began the 21st century at a disadvantage for Arabic due to ingrained French use, which helps explain why today English is making inroads as well (they had a bilingual environment already). It acknowledges that the Gulf states historically had strong Arabic environments, but in a single generation have imported massive foreign workforces and built English-centric sectors, which is an unprecedented historical experiment. The Historical dimension thus acts as a weighting/context layer: it might moderate the interpretation of S and P scores. A low Socioeconomic score in a historically secure Arabic heartland country might be more alarming (showing a new rapid decline) than a similarly low score in a historically linguistically mixed country (where it could be a continuation of legacy issues). Our index formula (as described in the Appendix) allows for such interplay by not simply averaging S, P, and H but considering H as a sort of baseline vulnerability multiplier.

## Composite Index and Early-Warning Thresholds

Each country's data in the above indicators is normalised (typically on a 0–10 or 0–100 scale for each indicator, where higher = better for Arabic vitality). Within each dimension, indicators are combined (sometimes weighted by data quality or presumed importance – e.g., we might weight "education medium" higher than "patents" in the Socioeconomic dimension). This yields three sub-scores (S, P, H). These then combine into an overall SPH-LENS index score, which we scale from 0 to 100 for convenience, where 100

would mean Arabic is completely secure (dominant in all domains, no competition) and 0 would mean extreme attrition (Arabic entirely replaced in most functions, though such a case doesn't exist in our dataset).

The index is meant for comparative and monitoring use, not an absolute measure of risk. We propose thresholds such that if a country's score falls below a certain value, it is flagged as at-risk. For instance, a score below 60 might indicate notable attrition requiring policy attention; below 40 a critical situation. These thresholds are determined by examining the distribution of scores and identifying natural breakpoints (detailed in Appendix). In our current analysis (hypothetical numbers for illustration), perhaps we find Gulf states scoring around 70 (due to strong policies but growing English use), Maghreb states around 50–60 (due to legacy French but efforts to pivot to English while Arabic is still not recovered fully), and some like Somalia or Djibouti maybe near 40 (Arabic being secondary for many). A country like Egypt might score relatively higher (it has a large Arabic media and education base, though facing English incursion) – say 75. If in a future measurement we see Egypt drop to 65, that delta signals a significant negative trend, even if 65 isn't "critical" in absolute terms. Thus, tracking change over time is as important as the score itself. A rapid decline in S or P subscores would ring an early alarm.

## Scope and limitations of SPH-LENS

It is important to emphasise that, in its current form, SPH LENS is an exploratory framework rather than a fully validated predictive model. The index is specified and illustrated using plausible indicator choices and country examples, but the article does not test its performance systematically against independent measures of political instability, social cohesion, or conflict. Data limitations and space constraints preclude, for instance, constructing time series SPH LENS scores for all Arab states and statistically assessing whether low or declining scores precede observable episodes of unrest. Our more modest contribution is to demonstrate the internal coherence and operational feasibility of the framework and to propose it as a structured early warning heuristic. Future work should refine the indicators, experiment with alternative weightings, and evaluate SPH LENS against other risk metrics before any strong claims about its forecasting accuracy are made.

## Comparative Insights: SPH-LENS vs. Traditional Models

Having established the SPH-LENS framework, the paper will examine how its application yields insights that differ from conventional language vitality assessments. We do so through brief case studies of four Arab countries – Tunisia, the United Arab Emirates, Morocco, and Egypt – chosen for their diverse linguistic landscapes and policy environments. For each, we highlight what SPH-LENS indicates about Arabic attrition risk and contrast that with their classification under traditional models like UNESCO or EGIDS. This comparative analysis underscores the added value of SPH-LENS in detecting subtle or emerging threats that a static classification might overlook.

### Tunisia

Arabic is the sole official language of Tunisia, spoken as a mother tongue by virtually the entire population (save

for a small francophone elite). By EGIDS criteria, Tunisian Arabic is "vigorous" (transmitted to children), and Modern Standard Arabic is secure as a national language. UNESCO's Atlas lists no threat.

### SPH-LENS Findings

Tunisia's Socioeconomic score reveals significant vulnerability. While public primary education is in Arabic, our indicators show that by secondary school and certainly in higher education, French (and increasingly English) dominate in science and technical fields. Recent reforms have introduced English from earlier grades, aiming to eventually use English in STEM, reflecting a sentiment that French – and by extension Arabic – are insufficient for globalisation.

SPH-LENS captures this through a declining "education in Arabic" indicator and a rising English proficiency among youth. Indeed, Tunisia topped the Maghreb in EF's English Proficiency Index 2021, and surveys show

a strong youth desire to pivot to English as the foreign language of choice.<sup>16</sup> Politically, Tunisia has no laws enforcing Arabic in the private sector or higher education; the constitution (2014) confirms Arabic as official but also promotes French by practice. Our framework thus notes a gap between official status and actual use in elite domains – something EGIDS would miss entirely.

Historically, Tunisia had one of the earliest and most pervasive French influences (a French protectorate from 1881 to 1956). Although post-independence Arabisation made strides (e.g., most primary schooling in Arabic by the 1980s), the entrenched role of French persisted in business and academia. Now the pivot to English could further marginalise Arabic academically if not managed.

### SPH-LENS Index

Tunisia scores moderate risk (~55/100 in our hypothetical composite), with a particularly low subscore in the Socioeconomic domain due to heavy foreign language penetration in education and web content. Traditional models would label Tunisia simply “safe (Level 1)” because virtually all Tunisians speak Arabic natively. SPH-LENS reveals the dynamic threat: that Tunisia might become a nation of Arabic speakers who increasingly think, learn, and work in another language. The early warning signs – youth attitudes, policy shifts – are visible now, not decades from now when perhaps intergenerational Arabic transmission could falter if the trend continued.

### United Arab Emirates (UAE)

Arabic is official; though a minority (Emirati nationals are less than 15% of the population), Arabic is still the home language for citizens. EGIDS might classify the local Gulf Arabic dialect as vigorous within its community, and Modern Standard Arabic is maintained via education/media. **SPH-LENS Findings**

The UAE illustrates a case where rapid modernisation and globalisation create extreme pressure on Arabic despite strong official support. Our Socioeconomic indicators show the UAE has one of the lowest shares of Arabic content on locally hosted websites (English and other languages dominate the UAE’s internet presence due to its expat-driven economy).

English is the lingua franca in private business, higher education, and even in many government offices (especially in Dubai, where expats vastly outnumber locals). A blog commentary bluntly noted the “propensity of Arabs [in the UAE] to speak English even when addressed in Arabic”. This is backed by anecdotal evidence: Emirati youth often code-switch heavily, and English-medium private schooling is common. Yet, politically, the UAE has undertaken high-profile initiatives to bolster Arabic: e.g., the government sponsors Arabic language conferences, has an Arabic requirement in some government communications, and the leadership frequently emphasises Arabic as core to identity. The Political dimension score for the UAE is mixed – positive for explicit language promotion policies (like Abu Dhabi’s law requiring Arabic signage, etc.), but negative for the structural reality that English is essentially the working language of much of the economy.

Historically, the UAE (and other Gulf states like Qatar and Bahrain) did not have a colonial legacy of a European language; their shift to English happened in a single generation due to oil, global business, and imported labour. SPH-LENS flags this as a historically acute shift: our Historical indicator shows “core Arabic region with late but fast foreign language influx”, which magnifies the risk because the population hadn’t been bilingual – the change is abrupt.

### SPH-LENS Index

The SPH-LENS composite for the UAE might be around 60/100 – slightly better than Tunisia perhaps in that the government is aware and trying to address it (and intergenerational transmission among citizens is still intact). But certain indicators blink red: for instance, university education in the UAE is overwhelmingly in English (except in a few programmes), and media consumption among the cosmopolitan populace skews to English (with major English newspapers and a large share of top social media influencers using English).

Traditional frameworks would not capture these vulnerabilities at all – the UAE would appear linguistically “safe” as long as Emirati households speak Arabic. But SPH-LENS reveals that Arabic’s functional load is shrinking in the UAE’s unique demographic setting.

An SPH-LENS analysis of corporate branding found that in the UAE’s dynamic private sector (free zones, startups), English names and communication are nearly universal, indicating low linguistic capital for Arabic in commerce. This suggests a possible future scenario where young Emiratis, though fluent in Gulf Arabic socially, might conduct most professional activities in English, making Arabic increasingly relegated to cultural and family contexts – a pattern akin to diglossia hardening into a functional split

## Morocco

Morocco is often cited as a bilingual country (Arabic and French), but Arabic (Darija dialect) is the mother tongue of the majority, and Modern Standard Arabic is official alongside Amazigh (since 2011). Ethnologue’s country entry would list Arabic as “National” (EGIDS 1) and not endangered, with French as a prestigious second language.

### SPH-LENS Findings

Morocco’s profile in our framework is one of transitional risk. After decades of French dominance in high domains, Morocco has recently seen an upswell in favour of English among youth, as well as continued diglossic challenges between Darija and Standard Arabic.

Socioeconomically, Morocco scores low: our indicators show a small proportion of scientific output in Arabic (most scholars publish in French or English), low Arabic web content relative to French (Morocco has a robust Francophone online scene and now an emerging Anglophone one), and strong English-learning trends (40% of young Moroccans choose English as the most important language to learn, versus 10% French). The EF EPI rank for Morocco is still low but improving, and dozens of private English-medium schools have opened.

Politically, Morocco has a nuanced story: Arabic is official, but the state never fully eliminated French from administration or higher education. In 2019, a controversial law brought back French for teaching science in high school, essentially admitting that the Arabisation policy of the 1980s had shortcomings. Simultaneously, the king established an institute for the Amazigh language, making Morocco officially multilingual (Arabic’s sole dominance is diluted). There is also a grassroots movement pushing for English to replace French as the primary foreign language taught. SPH-LENS reflects this fluid situation: Morocco’s Political dimension score is moderate – it has identity and language planning (the 2011 constitution gave Arabic and Amazigh official status, and there are efforts to regulate language in some settings), but also conflicting signals (e.g., switching instruction language indicates a lack of long-term strategy).

Historically, Morocco had a deep classical Arabic heritage but also one of the longest French occupations in the Arab world, leaving a population that to this day often finds French or Moroccan Arabic easier than formal Arabic for technical discussion. Our Historical dimension flags Morocco as a high vulnerability baseline (due to colonial legacy and diglossia).

### SPH-LENS Index

The SPH-LENS composite for Morocco might be around fifty-something out of 100, on par with Tunisia or slightly higher, depending on how the new English trend is interpreted (if English starts displacing French without strengthening Arabic, it’s just a new threat vector). Traditional indexes would likely call Morocco’s Arabic “safe” at EGIDS 1, with maybe a note about French influence but nothing alarming. SPH-LENS, by contrast, identifies Morocco as a borderline case where immediate attention is needed. Indeed, Morocco in our framework lights up “yellow” if not “red” on multiple indicators: the youth attitude shift is an early warning that another generation might value Arabic even less than the current one. On the positive side, SPH-LENS also notes Morocco’s potential: a large Arabic-speaking populace with strong Arabic literary output historically. If the state harnesses that (for example, by developing quality Arabic STEM education materials

rather than defaulting to French/English), the trajectory could yet improve. This exemplifies how our system isn't deterministic; it points to both threats and the latent capacity for revitalisation if policy intervenes.

## Egypt

Arabic in Egypt is extremely robust by numbers – over 100 million native speakers. Arabic is the sole official language and is used in all schooling (with English taught as a secondary language widely). EGIDS would rate Egyptian Arabic as vigorous (not endangered at all), and Egypt is often thought of as an “Arabic fortress” because of its role in Arab media (the Egyptian dialect dominates regional film/TV).

### SPH-LENS Findings

Egypt indeed fares better in our index than the above countries, but it is not without concerns. Socioeconomic indicators. On the plus side, a larger share of Egypt's higher education remains Arabic-medium (especially in social sciences, humanities, and K-12 education). However, certain fields (medicine, engineering) and prestigious universities (American University in Cairo, for example) operate in English, and there's a trend of affluent families gravitating to English-medium private schools. Our data might show, for instance, a slowly increasing number of international schools in Cairo and Alexandria and high English proficiency among the upper-middle class (Egypt's EF EPI rank is “low” overall, but urban elites score much higher). Egypt produces a substantial amount of Arabic content – it's a top producer of Arabic films, TV, and journalism – which boosts its digital and media indicators for Arabic. In W3Techs terms, the .eg domain likely has more Arabic content than, say, .ma or .tn. Yet, Egypt's youth are also part of the globalised internet culture: many follow English-language social media, and some prefer Western entertainment. SPH-LENS would mark

Political dimension for Egypt is relatively strong: Arabic is firmly enforced in government and public signage, and as of now there's no serious movement to replace it in education (though there are calls to improve English teaching for competitiveness). The constitution and laws back Arabic's primacy, and Egypt historically championed pan-Arab media (e.g., Sawt al-Arab radio in the 1950s, etc.).

The Historical dimension gives Egypt an advantage: minimal colonial linguistic impact (the British ruled but did not replace Arabic administratively; French influence was earlier and limited to elites in the 19<sup>th</sup> century). Thus, Egyptians have had uninterrupted transmission of Arabic classical and colloquial traditions.

### SPH-LENS Index

All this yields a higher SPH-LENS score, perhaps around 70–75/100, indicating relatively lower immediate risk. However, SPH-LENS still catches early warnings: for example, the brain drain of Egyptian professionals to the Gulf or West means a lot of higher education is happening in English (whether in Egypt's own universities that use English textbooks or abroad). Also, the booming private sector in tech and startups in Egypt often leans on English for coding and business plans, suggesting English competence is a gate to success even if Arabic is spoken informally. Another subtle indicator: OpenAlex data might show that while Egypt's volume of research is large, the portion in Arabic journals is small (Egyptian scholars often publish in English to reach international journals). A UNESCO study noted that Arabic is underutilised in scientific literature relative to Egypt's potential. Traditional indices wouldn't reflect that discrepancy. Our framework thus encourages Egyptian policymakers not to be complacent: even in the heart of the Arab world, the forces of English-centric globalisation are present. The difference is Egypt has the demographic and cultural heft to resist if it chooses – SPH-LENS underscores that it should make that choice deliberately (through, say, strengthening Arabic technical education and incentivising Arabic scientific publication) to avoid sliding down the path observed in smaller states.

In these case studies, we see that SPH-LENS provides a differentiated risk assessment. Traditional models lump all these cases together as “Arabic not endangered”; at most, a narrative might mention dialect vs. standard or the presence of a second language, but there is no numerical index for vitality threats when a language is still populous. SPH-LENS, by contrast, distinguishes that Morocco and Tunisia currently face more acute attrition dynamics than Egypt and that the UAE – despite wealth and official support – has a unique

challenge of demographics and English hegemony. Notably, SPH-LENS also highlights something counter-intuitive: a country with a very high EGIDS rating like the UAE (where Arabic is official and the child language of nationals) can in practice have a weaker outlook than a country with a lower EGIDS rating like Somalia (where Arabic is a second language for many). For instance, a hypothetical SPH-LENS score for Somalia might show moderate risk (say 50) largely because not everyone speaks Arabic to begin with, yet those who do maintain it in the religious domain, and Somalia's low global integration means English hasn't fully displaced local languages either. The UAE might score similarly or worse due to the sheer scale of English usage. This is an insight only gained by looking at functional and societal data, not just speaker counts.

Another insight is the identification of domain-specific vulnerabilities. SPH-LENS can tell us, for example, that Tunisia's main issue lies in higher education and research (S dimension), whereas the UAE's is in everyday business and population makeup (mix of S and H factors), and Morocco's is in a tug-of-war in education policy (P dimension) combined with historical baggage (H). This granularity allows tailored solutions: Tunisia might focus on enriching Arabic scientific vocabulary and confidence; the UAE on incentivising private sector Arabic use and content creation; Morocco on stabilising language policy (perhaps using Arabic for lower levels, a controlled switch to English in higher ed but accompanied by strong Arabic reinforcement, etc.); and Egypt on leveraging its cultural industries to make Arabic "cool" and ensuring economic value for Arabic skills.

## Policy Implications

Framing Arabic language vitality as a national security concern demands a fundamentally different policy approach than treating it as a cultural heritage issue. The findings from the SPH-LENS analysis underscore that governments must act early and strategically to stem and reverse the attrition of Arabic. In this section, we outline policy implications of our framework and propose recommendations tailored to the SPH dimensions, aligning with the overarching goal of an early-warning system. The intent is to guide policymakers on how to operationalise SPH-LENS: not only monitoring the indicators but also responding with informed interventions. We maintain the academic tone but direct the focus to practical measures, as the ultimate purpose of developing this framework is to enable replicable, transparent, and effective action.

### Institutionalize SPH-LENS Monitoring

First and foremost, Arab governments (perhaps coordinated through the Arab League or ALECSO) should institutionalise the kind of monitoring we have presented. This could mean establishing a "Language Vitality Observatory" that regularly collects data on the agreed indicators – from education ministry stats to tech surveys – and publishes an annual Arabic Language Vitality Report. Just as central banks watch economic indices to adjust policy, cultural and education ministries (in collaboration with national security councils) should watch linguistic indices. The Appendix of this paper provides a ready blueprint for indicators and scoring; states can refine it to their context. The key is transparency: by using existing data sources (UNESCO, World Bank, etc.), the results can be trusted and compared year to year. We recommend that thresholds be set such that, if a country's score falls or approaches the "red zone", an inter-ministerial task force is triggered to investigate causes and recommend remediation. This is analogous to how a public health early warning might work for an epidemic. For example, if next year's SPH-LENS index for Tunisia drops by 5 points due to a sudden surge in English-medium programmes, that should prompt a policy review on language in education.

### Revamp Language Education Policies

Education is the front line of either losing or saving Arabic's future status. SPH-LENS data often highlights a mismatch: parents and students flock to foreign-language education for perceived economic benefits. To change this calculus, governments must raise the socio-economic capital of Arabic. One approach is bilingual advantage without Arabic abandonment: implement robust Arabic-English (or Arabic-French) bilingual education models, where Arabic is used to teach certain subjects (especially social studies, religious education, Arabic language arts, etc.) at high standards, while foreign languages are taught deeply as languages and used in selected technical subjects. This prevents the "either-or" scenario where choosing modernisation means dropping Arabic. Countries like the UAE and Qatar could require that even in branch campuses of foreign

universities, a certain proportion of courses on local culture or history be in Arabic to maintain academic Arabic usage. Translation and curriculum development are critical: invest in translating world-class science and math textbooks into Arabic (as was done in Syria and Iraq historically) so that studying these in Arabic doesn't mean using outdated materials. A positive example is the King Abdullah initiative in Saudi Arabia that translated hundreds of top textbooks into Arabic for universities – such efforts must continue and expand regionally. Additionally, terminology modernisation must be well-funded: organisations like Arabic language academies should coin and disseminate Arabic technical terms (for IT, medicine, etc.) so that lecturers and media have the vocabulary to discuss cutting-edge topics in Arabic, rather than defaulting to English terms. This directly tackles the “prestige barrier” where Arabic is seen as not having the words for modern concepts.

### **Economic Incentives for Arabic Proficiency**

Governments should consider incentives that make proficiency in Arabic economically rewarding, reversing the current incentive structure. This could include hiring preferences or bonuses for civil servants and employees in public companies who demonstrate strong Arabic (e.g., through certification), while also requiring a baseline of Arabic ability in foreign companies operating in the country (for instance, requiring that customer service by telecom and utility companies be available in fluent Arabic – many already do, but it should be regulated and enforced). Another idea is to encourage Arabic content creation industries – perhaps through grants or tax breaks – whether it's Arabic software development, Arabic media startups, or Arabic academic journals (e.g., funding an “Arabic Science Journal” of high quality and incentivising researchers to publish in it with recognition and rewards). If a university professor gets as much credit for publishing in a top Arabic journal as in an international English journal (perhaps through local grant systems), they might invest more in Arabic research output. On a larger scale, the Arab League could develop an Arab Knowledge Portal translating global research (akin to the 1000 Books initiative of the UAE) and hosting Arabic papers to elevate the status of Arabic as a language of knowledge, not just heritage.

### **Strengthen Language Governance**

The Political dimension in SPH-LENS indicates that policy coherence is often lacking. We recommend that each Arab state formulate a National Language Strategy – a document akin to a security strategy – that lays out where and how Arabic should be strengthened and how foreign languages will be taught without displacing Arabic. Some specific measures: enforce existing official language laws (if Arabic is official, laws and regulations should stipulate its use in all governmental communication; Morocco, for instance, could ensure all government websites which are now often bilingual maintain complete Arabic versions). Where constitutions are being updated or amended (as several Arab countries have in the past decade), they enshrine provisions that commit the state to protect and promote Arabic in the face of globalisation. Consider adopting a form of the “Toubon Law” (like France uses to protect French) – an Arab-world version could mandate Arabic usage in certain domains (advertising, public signage, product labelling, etc.) while not banning other languages but ensuring Arabic visibility and primacy. Another governance aspect is to equip language academies and cultural bodies with more authority and funding – these should not be just symbolic. For instance, the Academies in Cairo, Damascus, etc., could coordinate to produce unified school lexicons, and ministries and governments should heed their recommendations on avoiding unnecessary foreign jargon in official usage.

### **Recognise Multilingualism as a Resource, Not Only a Risk**

A security-oriented framing of Arabic vitality must avoid collapsing into a zero-sum view in which other languages are treated simply as threats. English, French, Amazigh, Kurdish, and other languages present across the Arab world also generate valuable human capital and external linkages. The challenge, from an SPH-LENS perspective, is to prevent these languages from monopolising access to high-status domains while Arabic is confined to low-capital or purely symbolic functions. Policy should therefore aim to design robust bilingual and multilingual arrangements – for example, high-quality Arabic-medium schooling combined with strong foreign-language instruction, or firm guarantees for minority languages within an Arabic-dominant public sphere – so that linguistic diversity becomes an asset for resilience and outreach, even as Arabic remains the principal integrative and legitimising language of state and society.

## Cultural Campaigns to Rebrand Arabic

Public perception matters hugely. If youth view Arabic as “dated” or only for informal use, no policy will stick unless that mindset shifts. Governments, media, and civil society can collaborate on campaigns to enhance the image of Arabic. For example, the UAE in recent years has had annual initiatives around World Arabic Language Day, showcasing Arabic’s contributions to science and arts. Such efforts should be expanded: broadcasting content about Arabic’s global contributions (reminding that Arabic was once the lingua franca of science), highlighting successful Arab scientists and entrepreneurs who champion Arabic, and leveraging influencers – e.g., popular Arab YouTubers or musicians – to incorporate and valorise Arabic in their content. The goal is to break the association that using English is “cooler” or more future-orientated. Part of this is also modernising Arabic teaching: making learning formal Arabic more engaging in schools so that students don’t see it as a tedious subject disconnected from their lives. Initiatives like developing Arabic educational video games, AI assistants that work in Arabic dialects and standard, and ensuring Arabic support in emerging tech (so that, say, when kids code, they can use Arabic in comments or logic if they want) can all help root the language in modernity.

## Regional Cooperation and Knowledge Sharing

Language planning is more effective when done collectively, especially for a language common to many states. The Arab League and organisations like ALECSO and ISESCO should take SPH-LENS as a cue to intensify cooperation. This could mean a regional observatory mentioned earlier, but also collaborative projects like a Pan-Arab Digital Library where content from all countries is shared (this exists partially in initiatives like the “Arab Digital Content” project) or an Arab Languages Early-Warning Network where countries share best practices and possibly coordinate policies to avoid working at cross-purposes. For example, if Morocco finds success in shifting some of its tertiary programmes from French to Arabic (hypothetically via translation and teacher training), it should share that model with Algeria or Tunisia. Conversely, Gulf states that have managed to keep Arabic as the medium in certain sectors (like Saudi Arabia in general education) can exchange experiences with countries struggling in that area. There could also be a joint fund to support Arabic localisation of technology – like supporting Arabic NLP (natural language processing) research so that Arabic voice recognition, OCR, and machine translation are as good as English. This reduces the barrier of using Arabic in digital spaces.

## Leverage Arabic for Soft Power

Governments often talk about projecting soft power through culture; Arabic should be a core element of that. Just as France has Alliance Française and China has Confucius Institutes, Arab countries (perhaps under the Arab League flag) could expand Arabic language and culture centres worldwide. This not only benefits non-Arabs learning Arabic but also boosts the prestige of Arabic globally, which in turn can increase local pride. Imagine if Arabic were more commonly taught as a foreign language abroad – Arabs might feel their language is indeed international and valuable. Already, Arabic is one of the UN’s official languages and used diplomatically; building on that, more academic partnerships can be fostered where Arabic is a medium – e.g., encourage reputable foreign universities to offer programmes in Arabic or joint degrees with Arab universities that include Arabic content. Historically, Arabic thrived when it was at the centre of a cosmopolitan world (the Abbasid era translation movement, etc.). Reviving a bit of that spirit – positioning Arabic as a language of global dialogue (for South-South cooperation, Islamic world scholarship, etc.) – can reinforce internal confidence in the language.

## Addressing Historical Inequities

In places with diglossia or multilingual populations, policies must be sensitive and inclusive. For example, in North Africa, promoting Arabic must go hand in hand with respecting Amazigh language rights (so that pro-Arabic doesn’t become seen as anti-identity for Amazigh communities). In Sudan or Somalia, promoting Arabic should consider integration with local languages rather than replacement to gain buy-in. This means bilingual education models and intercultural programmes. The positive effect is when Arabic is seen as an addition that empowers (connecting to a wider Arab identity and opportunities) rather than a top-down

imposition that replaces mother tongues, people embrace it more fully. This was a lesson from some failed Arabisation efforts that disregarded local languages and faced pushback. A more inclusive language planning will strengthen Arabic's role without breeding resentment that could later cause communities to abandon it. Essentially, treat Arabic as a core part of a multilingual national identity, not a monolithic one, in those contexts.

## Continuous Research and Feedback

Lastly, policymakers should treat SPH-LENS as a living framework. They should encourage local universities and think tanks to refine the indicators, study the efficacy of interventions, and research sociolinguistic trends (e.g., conduct surveys on language use in families or how new technologies like AI assistants might affect Arabic usage). The data gathered should feed back into policy adjustments. For example, if after a few years of an intervention (say, introducing more Arabic content in higher education) the SPH-LENS score doesn't improve or youth attitudes remain the same, investigate why – maybe the quality was an issue, or job market signals still undermined the effort. Because language use is ultimately a sum of individual choices responding to incentives, policy has to be flexible and responsive to what the indicators tell us about those choices.

## FINDINGS AND CONCLUSION

Arabic's status in the Arab world today sits at an inflection point. This paper has argued that what might appear, by traditional metrics, to be a secure language is, in fact, confronting a slow-burning crisis of attrition with profound socio-political ramifications. The paper introduced the SPH-LENS framework as both an analytical tool and a call to action: a comprehensive early-warning system that governments and researchers can use to detect and respond to Arabic language decline before it reaches a point of no return.

The findings of the SPH-LENS analysis are sobering. Across multiple Arab states, the paper identified concrete indicators of domain loss – from the lecture halls of universities to the boardrooms of new startups, from the smartphones of teenagers to the laboratories of scientists. In each sphere, we see English (and in some cases French) encroaching on territory once firmly held by Arabic. Crucially, these shifts are happening not as isolated anomalies but as part of a broader pattern driven by globalisation, economic incentives, and historical context. Traditional vitality indices, which still classify Arabic at the highest tiers of safety, fail to capture these dynamics. Our work demonstrates that a language can have over 400 million speakers and official status in dozens of countries and yet still be at genuine risk – not of extinction, but of strategic degradation, where it loses its ability to function as the backbone of a modern society.

The SPH-LENS framework's three dimensions – Socioeconomic, Political, and Historical – proved essential in disentangling the complexity of Arabic's situation. The Socioeconomic analysis revealed a self-reinforcing cycle: as English and other languages accrue greater prestige and utility (being seen as keys to better jobs, quality education, and global connectivity), individuals shift toward them, which in turn further devalues Arabic's capital. The Political examination uncovered that many Arab governments have inadvertently allowed or even facilitated a kind of linguistic laissez-faire, where global languages fill the void of modernisation, sometimes with policy endorsement (as in the re-Francophone-isation of Moroccan science education). We also saw instances of positive political will – policies and campaigns to bolster Arabic – but this need scaling up and better coordination. The Historical perspective reminded us that today's challenges are layered on top of yesterday's legacies: colonies where French or English took root face a different battle than heartland countries, and yet even the latter are not immune to rapid shifts given the right (or wrong) conditions.

One of the central implications of this study is the extent to which language vitality is entangled with wider questions of national security. The analysis offered through SPH-LENS is deliberately cautious: it maps plausible pathways through which stratified multilingualism and domain loss might interact with existing socioeconomic and political stresses, rather than asserting that changes in Arabic usage will mechanically produce instability. In this sense, the framework should be read as a structured early-warning heuristic – one that flags configurations of linguistic capital, policy, and historical vulnerability that merit closer empirical and policy attention – rather than as a calibrated forecasting device. Equally, the findings underscore that

multilingualism can coexist with, and even strengthen, societal resilience when language hierarchies are actively managed in inclusive and egalitarian ways.

In conclusion, the SPH-LENS framework reframes the narrative around “Arabs without Arabic” from one of lament to one of proactive response. It is a lens that allows us to see clearly the sociopolitical threads connecting language to stability. What we see through that lens is a mixture of warning and hope: warning, in the quantifiable signs of attrition already underway; hope, in that early detection opens the possibility for timely intervention. The Arab world stands at a juncture where deliberate, informed choices in policy and society will determine whether Arabic continues to serve as the vessel of a great civilisation’s legacy and future or gradually fades into a secondary role on its own soil. The findings and recommendations of this paper aim to support the former outcome. The Arabic language, with its millennia of rich history, need not be a casualty of globalisation. With vigilance, creativity, and commitment – guided by frameworks like SPH-LENS – Arabic can be revitalised as a living force that both honours its past and ignites the future of the Arab world.

## **Declaration of Funding**

No funding was received

## **Conflict of Interest**

No potential competing interest was reported by the authors

## **ACKNOWLEDGEMENT**

We would like to express our sincere gratitude to colleagues who provided invaluable guidance and constructive feedback throughout the development of this study. Special thanks are due to Al Habtoor Research Centre in Dubai (UAE), which offered access to resources and scholarly discussions, which enriched both the theoretical and empirical dimensions of this work. We extend our appreciation to our families for their patience, encouragement, and unwavering support during the research and writing process.

## **REFERENCES**

<sup>1</sup> Ayres, Alyssa. 2003. “The Politics of Language Policy in Pakistan.” In *Fighting Words: Language Policy and Ethnic Relations in Asia*, edited by Michael E. Brown and Sumit Ganguly. Cambridge, MA: MIT Press.

<sup>2</sup> Berets, Rachel. “The Latest in Language Confusion: Morocco Switches Back from Arabic to French.” Al-Fanar Media, May 2, 2020. Accessed November 2, 2025. <https://al-fanarmedia.org/2020/03/the-latest-in-language-confusion-morocco-switches-back-from-arabic-to-french/>.

<sup>3</sup> Rix, Mark. “Lost in Translation: Is the Arabic Language Being Eroded?” Salford Business School Blog. Accessed November 4, 2025. <https://blogs.salford.ac.uk/business-school/tag/lost-in-translation/>.

<sup>4</sup> DeVotta, Neil. 2003. “Ethnolinguistic Nationalism and Ethnic Conflict in Sri Lanka.” In *Fighting Words: Language Policy and Ethnic Relations in Asia*, edited by Michael E. Brown and Sumit Ganguly. Cambridge, MA: MIT Press.

<sup>5</sup> Greenberg, Robert D. 2004. *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration*. Oxford: Oxford University Press.

<sup>6</sup> Grin, François. 1999. “Language Policy in Multilingual Switzerland: Overview and Recent Developments.” ECMI Brief no. 2. Flensburg: European Centre for Minority Issues.

<sup>7</sup> Hayday, Matthew. 2005. *Bilingual Today, United Tomorrow: Official Languages in Education and Canadian Federalism*. Montreal and Kingston: McGill-Queen’s University Press.

<sup>8</sup> ALTA Language Services. “The UNESCO Atlas of the World’s Languages in Danger,” March 6, 2009. Accessed November 7, 2025. <https://unesdoc.unesco.org/ark:/48223/pf0000187026/PDF/187026eng.pdf.multi>

<sup>9</sup> Alshehri, F.A., Alshehri, F.A., Qassem, M. et al. The dynamics of translation from and into Arabic in the Arab

world: bibliometric analysis of the Index Translationum UNESCO database (1979–2012). *Humanit Soc Sci Commun* **12**, 43 (2025). <https://doi.org/10.1057/s41599-024-04225-5>

<sup>10</sup> United Nations Development Programme. *Arab Human Development Report 2003: Building a Knowledge Society*. New York: United Nations Development Programme, 2003. Accessed November 07, 2025. <https://hdr.undp.org/content/arab-human-development-report-2003>

<sup>11</sup> Aspadmin. “Regional Consultations launched for building Arabic Chapter of World Atlas of Languages.” UNESCO, April 23, 2021. Accessed November 6, 2025. <https://www.unesco.org/en/articles/regional-consultations-launched-building-arabic-chapter-world-atlas-languages>

<sup>12</sup> British Council, Shift to English in Morocco (April 2021), accessed November 27, 2025, [https://www.britishcouncil.ma/sites/default/files/shift\\_to\\_english\\_in\\_morocco\\_16042021\\_v2.pdf](https://www.britishcouncil.ma/sites/default/files/shift_to_english_in_morocco_16042021_v2.pdf)

<sup>13</sup> Asharq Al Awsat. “UNESCO: Arabic Is One of the Pillars of Human Civilization,” December 19, 2020. <https://english.aawsat.com/home/article/2691396/unesco-arabic-one-pillars-human-civilization>.

<sup>14</sup> Alshehri, F.A., Alshehri, F.A., Qassem, M. et al. The dynamics of translation from and into Arabic in the Arab world: bibliometric analysis of the Index Translationum UNESCO database (1979–2012). *Humanit Soc Sci Commun* **12**, 43 (2025). <https://doi.org/10.1057/s41599-024-04225-5>

<sup>15</sup> Pimienta, Daniel et al. “The method behind the unprecedented production of indicators of the presence of languages on the Internet.” *Frontiers in research metrics and analytics* vol. 8 1149347. 18 May. 2023, doi:10.3389/frma.2023.1149347

<sup>16</sup> EF Education First, EF English Proficiency Index 2021: A Ranking of 112 Countries and Regions by English Skills (2021), 5–6, <https://www.ef.com/assetscdn/.../ef-epi-2021-english.pdf>

## Appendix: SPH-LENS Scoring Methodology and Indicator Details

(This appendix provides a structured overview of the indicators, normalisation, weighting, and index formula used in the SPH-LENS framework. It is intended as a reference for replicating the study or implementing the framework in policy monitoring. Each section below corresponds to one aspect of the methodology).

### Indicator List by Dimension

#### Socioeconomic (S) Indicators

**Table** Error! Main Document Only.- Socioeconomic (S) Indicators of the SPH-LENS Framework

Code	Indicator	Definition	Justification
<b>S1</b>	<b>Education Medium Index</b>	Proportion of students in secondary and tertiary education taught primarily in Arabic. Data source: UNESCO UIS, national education stats. (0 = 0%, 10 = 100% Arabic-medium at those levels, linear scale).	Measures presence of Arabic in formal high-level knowledge transfer (a decline indicates domain loss).

S2	<b>Scientific Publication Language Share</b>	Percentage of research publications by authors from the country that are in Arabic (as opposed to English/French). Data: OpenAlex or Scopus analysis. (0 = 0%, 10 = $\geq 20\%$ in Arabic, with interpolation).	Direct gauge of Arabic usage in creating new knowledge.
S3	<b>Web Content in Arabic</b>	Percentage of websites or online content from the country in Arabic. Data: W3Techs (for country-specific if available) or alternative local internet surveys. (0 = 0%, 10 = $\geq 50\%$ content in Arabic).	Indicates digital vitality of Arabic; low values signal digital domain loss.
S4	<b>English Proficiency Index (inverse)</b>	EF EPI score of country, inverted because higher English often correlates with more shift from Arabic. We convert EF's 800-point scale to 0-10 or use rank-order inversion.	High English ability can be a risk factor (though also a development indicator; we contextualise rather than penalise it solely).
S5	<b>Business Language Indicator</b>	Proportion of top 100 domestic companies (or an index) that use Arabic in company name, official website, and internal comms. Data: corporate info. (Composite score 0-10 combining those sub-metrics).	Reflects market valuation of Arabic vs. foreign language in the corporate sphere.
S6	<b>Media Consumption Language</b>	Share of population consuming news/entertainment primarily in Arabic vs. other languages (from surveys). If unavailable, proxy by number of major foreign-language media outlets. (0 = majority consume foreign media, 10 = nearly all consume Arabic media).	Cultural penetration of Arabic in daily life.
S7	<b>Patent Filing Language</b>	Percentage of patents by residents filed in Arabic (for countries with local IP offices) or trademarks in Arabic. Data: WIPO or national IP office. (0 = 0%, 10 = $\geq 50\%$ ).	Niche but indicative of Arabic use at the highest innovation levels.

(Note: Each S indicator is normalised to 0-10. Weightings: Education (S1) and Web (S3) are given higher weight (e.g., 20% each), Scientific output (S2) 15%, English proficiency (S4) 10%, Business (S5) 15%, Media (S6) 10%, and Patent (S7) 10%. Weights based on perceived impact and data reliability. Total S score is the sum of weighted indicator scores).

### Political (P) Indicators

**Table** Error! Main Document Only.- Political (P) Indicators of the SPH-LENS Framework.

Code	Indicator	Definition	Justification
P1	<b>Official Status &amp; Constitutional Protection</b>	Coded 0-10 based on constitutional language provisions. 10 = Arabic is the sole official language, mandated in all state functions; 5 = Arabic co-official with another; 0 = Arabic not official. Data: Constitute Project, legal texts.	Baseline legal status.

P2	<b>Language Policy Enforcement</b>	Score 0-10 on presence/enforcement of laws (Arabic in signage, education language laws, civil service requirements, etc.). 10 = comprehensive language law fully enforced; 5 = some policies but patchy enforcement. Data: national legislation records.	Measures practical commitment.
P3	<b>Media and Press Freedom (language aspect)</b>	Combines Freedom House Press Freedom score with an assessment of language use. Qualitatively adjusted: a partially free media with strong local language content might score higher than a free media dominated by foreign content.	Accounts for influence of openness on language in media.
P4	<b>E-Government Language Accessibility</b>	Score 0-10 based on the extent key government digital services are available in Arabic. 10 = all services Arabic-first; 5 = bilingual, with English sometimes primary; 0 = mostly English. Data: government portals analysis, GTMI data.	Ensures digital governance includes Arabic.
P5	<b>Education Language Policy Consistency</b>	Examines whether official policy on medium of instruction aligns with practice. 10 = full alignment pro-Arabic; 0 = policy or practice effectively excludes Arabic. Data: Ministry of Education policies vs. actual curricula.	Identifies policy gaps/honesty.
P6	<b>International Cultural Influence</b>	Number of foreign cultural institutes/schools per capita (inverted score: more foreign schools = lower score). Normalised using range found in data.	Proxy for external hegemony presence.
P7	<b>Civil Society &amp; Initiatives for Arabic</b>	Count/score of active initiatives, NGOs, or campaigns promoting Arabic. 10 = vibrant movement (gov programs + grassroots); 0 = none.	Political will is not just gov – society support matters.

(P indicators normalised to 0-10, weighted somewhat evenly: Official status 15%, Policy enforcement 20%, Media freedom/usage 15%, E-gov 10%, Education policy 20%, Intl influence 10%, Civil society 10%. P score is a weighted sum).

### Historical (H) Indicators

**Table Error! Main Document Only.- Historical (H) Indicators of the SPH-LENS Framework.**

Code	Indicator	Definition	Justification
H1	<b>Colonial Legacy Impact</b>	Categorical scoring based on depth of colonial language imposition. 10 = no significant colonial language legacy; 5 = long rule but successful Arabization; 0 = Arabic introduced recently or not native.	Baseline vulnerability from history.
H2	<b>Indigenous Multilingualism</b>	Score 10 if the population >90% speaks Arabic natively; lower scores for countries with significant communities speaking other languages. Data: Ethnologue, census.	A higher score means Arabic is historically the mother tongue of nearly all, giving it resilience.

<b>H3</b>	<b>Historical Use in Elite Domains</b>	Retrospective look at the last ~100 years to determine if Arabic was the language of high culture and education. 10 = yes; 5 = partly; 0 = no.	Accounts for past status momentum.
<b>H4</b>	<b>Arabization Policy History</b>	10 = early and thorough Arabization (e.g., Syria); 5 = attempted late/partial (Maghreb); 0 = never.	Captures efforts to reverse legacy; success implies the current older professional generation is comfortable in Arabic.
<b>H5</b>	<b>Continuity of Arabic Education Tradition</b>	Presence of long-standing Arabic universities or religious institutions. 10 = continuous (e.g., Al-Azhar); 0 = none historically.	Provides depth of intellectual reservoir in Arabic.
<b>H6</b>	<b>Prior Language Shift Episodes</b>	Measures stability of language adherence. 10 = no major shifts; 5 = one significant oscillation; 0 = multiple/ongoing shifts away from Arabic in key domains.	Indicates stability or volatility of language policy over time.
<b>H7</b>	<b>Geographic Peripherality</b>	Measures distance from the "core" Arab world. 10 = core (Arabian Peninsula, Nile); 5 = intermediate (Levant); 0 = far periphery. Can be mapped by distance from Mecca or Cairo.	Accounts for historical isolation from Arabic cultural centres.

(H indicators are a mix of qualitative and quantitative. Weighting: Colonial legacy 20%, Indigenous multilingual 20%, Elite domain history 15%, Arabisation policy 15%, Education tradition 10%, Shift episodes 10%, Peripherality 10%. H score from weighted sum).

### Normalization and Scoring

Each raw indicator is converted to a 0-10 scale as per the descriptions above. For quantitative ones, we use min-max normalisation based on logical bounds or observed range across Arab states. For example, if the lowest web Arabic content is 0.5% and the highest is 98%, we set 0→0.5% and 10→98%. In practice, we often have natural bounds (0-100%). Qualitative indicators (like H1 colonial impact categories) were pre-assigned scores.

After normalisation, we applied weights to combine indicators into S, P, and H dimension scores (0-10 or 0-100 if scaling differently). We scaled each dimension to 0-100 for ease but internally kept careful weight proportions.

Finally, the SPH-LENS composite index is calculated. We considered two approaches: a simple weighted average of S, P, and H or a multiplicative approach treating H as a multiplier (since historical factors condition the other two). We opted for a weighted average to maintain transparency:

$$\text{Index} = w_S * S_{\text{score}} + w_P * P_{\text{score}} + w_H * H_{\text{score}}$$

$H_{\text{score}}$ , with weights summing to 1. In our prototype, we gave slightly higher weight to S and P (because they reflect current dynamics) and slightly lower to H (which is more static). E.g.,  $w_S = 0.4$ ,  $w_P = 0.4$ ,  $w_H = 0.2$ . This means the index is 40% determined by socioeconomic indicators, 40% by political factors, and 20% by the historical baseline. We chose this after testing that historical scores, which don't change quickly, should not overwhelm the index but still factor in. A country can thus overcome a poor historical baseline with strong present actions (and vice versa; a good legacy can be squandered with poor current policies).

## Threat Level Thresholds

Based on the composite index (0-100 scale), we define:

**Table** Error! Main Document Only.- Threat Level Thresholds of the SPH-LENS Framework.

Threat Level	Score Range	Status Description	Implications & Examples
<b>Green (Secure)</b>	<b>70 – 100</b>	Arabic is thriving in most domains, with minimal signs of attrition.	<b>Implication:</b> The language is secure, though trending indicators should still be watched.  <b>Example:</b> Egypt (scoring ~75) would be on the lower end of this zone.
<b>Yellow (Watchful)</b>	<b>50 – 69</b>	Noticeable attrition in some areas; serves as an early warning.	<b>Implication:</b> Immediate policy action is needed to prevent worsening conditions.  <b>Example:</b> Morocco (estimated ~55–60) likely falls here.
<b>Orange (At Risk)</b>	<b>30 – 49</b>	Significant attrition is present. If unaddressed, it could lead to a major decline in prestige and use within a generation.	<b>Implication:</b> High risk of losing functional domains.  <b>Example:</b> Lebanon or Tunisia might fall in the high 40s or low 50s.
<b>Red (Critical)</b>	<b>&lt; 30</b>	Arabic is effectively marginal in key domains.	<b>Implication:</b> Danger of intergenerational transmission failure or heavy societal splits.  <b>Example:</b> Potentially small states like Djibouti if youth usage remains very limited.

These thresholds were informed by the range of scores we computed. None of the 22 Arab states scored in the Red in our initial run (which is good – there is still time to correct course), but a few flirt with Orange. The majority were Yellow, meaning action is needed to push them back to Green. These thresholds can be refined as more data accumulates.

### Example Index Calculation (Hypothetical Country X)

To illustrate, suppose Country X has: - S dimension: Education in Arabic moderate (score 6), Sci publications low (2), Web content 4, English proficiency moderate (5), Business use 5, Media use 6, Patent 1. Weighted sum maybe = 5.0 (50/100). - P dimension: Official status strong (10), Policy enforces 7, Press freedom moderate but local content low (5), E-gov bilingual (5), Education policy consistent (8), Foreign schools many (3), Civil society active (7). Sum ~6.5 (65/100). - H dimension: Colonial legacy heavy (3), Indigenous bilingual (7), Elite history partial (5), Arabisation strong (8), Edu tradition strong (8), past shifts none (10), periphery intermediate (5). Sum ~6.0 (60/100).

Composite = 0.450 + 0.465 + 0.2\*60 = 20 + 26 + 12 = 58. Country X would be a yellow zone – needing watchful attention.

## Data Sources Summary

For transparency, below is a summary of key data sources and references used to populate SPH-LENS indicators in our study:

**Table Error! Main Document Only.- Data Sources for the SPH-LENS Framework.**

Source	Usage & Metric
<b>UNESCO Institute for Statistics (UIS)</b>	Provided data on language of instruction, enrollment figures, and literacy rates.
<b>Ethnologue / World Atlas of Languages</b>	Used for data on language status and speaker populations.
<b>W3Techs Web Technology Surveys</b>	Source for web content language statistics.
<b>Wikimedia Statistics</b>	Used (where available) to track contributions to Arabic Wikipedia.
<b>EF English Proficiency Index Reports</b>	Provided country scores and trends regarding English proficiency.
<b>OpenAlex Database</b>	Used for bibliometric analysis of scientific publication languages.
<b>WIPO &amp; World Bank</b>	Data for patent and innovation metrics, as well as GTMI data for e-government context.
<b>Constitute Project</b>	Source for analyzing constitutional language clauses and official status.
<b>Freedom House &amp; Reporters Without Borders</b>	Data on press and internet freedom, contextualized within the language environment.
<b>ALECSO / Arab League Reports</b>	Reports such as the “State of Arabic Language” used for qualitative inputs.
<b>Academic &amp; News Sources</b>	Specific case evidence (e.g., Nature studies, British Council surveys, local media) used to score specific indicators.

All sources are cited in context in the main text where their data is discussed, per the citation style (e.g., for Morocco’s law change, etc.). By compiling these diverse data into one framework, SPH-LENS ensures a multi-faceted yet unified assessment. Users of this framework in the future should update the data from these sources (many are annual or regularly updated) to keep the index current.

## Limitations and Future Refinements

The paper acknowledges certain limitations in our methodology:

**Table Error! Main Document Only.- Limitations and Future Refinements of the SPH-LENS Framework.**

Limitation	Description	Future Refinement Strategy
<b>Data Gaps</b>	Comprehensive data is not available for every country (e.g., precise publication language breakdowns). Regional proxies were sometimes used.	Users should update data sources regularly as they become available to improve precision.

<b>Qualitative Scoring</b>	Certain Political and Historical indicators rely on expert judgment, which introduces subjectivity.	Expand quantifiable measures (e.g., indexing actual counts of foreign schools) to improve objectivity.
<b>Dialects vs. Standard Arabic</b>	The framework treats "Arabic" monolithically, missing the nuance of diglossia where dialects thrive but Standard Arabic weakens.	Future versions could add indicators for the "diglossic gap" or dialect use in media to better assess the vitality of the standard form.
<b>Dynamic Interactions</b>	The current weighting scheme is linear, whereas language shift may experience non-linear "threshold effects".	Future modeling could explore composite functions that are not purely linear to capture accelerated decline.
<b>Language Attitudes</b>	Attitudes are currently proxied via behaviors (e.g., enrollment, media use) rather than direct sentiment.	Incorporate direct survey data (e.g., youth preference for reading in English vs. Arabic) to strengthen the Socioeconomic dimension.
<b>Validation Needs</b>	The framework is theoretical and requires testing against real-world outcomes over time.	Researchers are encouraged to conduct longitudinal validation to see if low scores predict actual social cleavage or language loss in the next generation.

Despite these, we believe the framework is robust enough to be actionable. Refinements can be made by the community; we encourage other researchers to tweak weights or add indicators and test the index against real-world outcomes (e.g., do countries with lower SPH-LENS scores subsequently show more evidence of social cleavage or higher use of foreign languages among the next generation? Such longitudinal validation would be valuable).