Chapter 3
Digital Language Equality: Definition, Metric, Dashboard

Federico Gaspari, Annika Grützner-Zahn, Georg Rehm, Owen Gallagher, Maria Giagkou, Stelios Piperidis, and Andy Way

Abstract This chapter presents the concept of Digital Language Equality (DLE) that was at the heart of the European Language Equality (ELE) initiative, and describes the DLE Metric, which includes technological factors (TFs) and contextual factors (CFs): the former concern the availability of Language Resources and Technologies (LRTs) for the languages of Europe, based on the data included in the European Language Grid (ELG) catalogue, while the latter reflect the broader socio-economic contexts and ecosystems of the languages, as these determine the potential for LRT development. The chapter discusses related work, presents the DLE definition and describes how it was implemented through the DLE Metric, explaining how the TFs and CFs were quantified. The resulting scores of the DLE Metric for Europe's languages can be visualised and compared through the interactive DLE dashboard, to monitor the progress towards DLE in Europe.[1]

1 Introduction and Background

The META-NET White Paper Series (Rehm and Uszkoreit 2012) showed the clear imbalance in terms of technology support for 31 European languages as of 2012 (see Chapter 1). Beyond the official European and national languages, more than 60 regional and minority languages (RMLs) are protected by the European Charter for Regional or Minority Languages and the Charter of Fundamental Rights of

Federico Gaspari · Owen Gallagher · Andy Way
Dublin City University, ADAPT Centre, Ireland, federico.gaspari@adaptcentre.ie, owen.gallagher@adaptcentre.ie, andy.way@adaptcentre.ie

Annika Grützner-Zahn · Georg Rehm
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Germany, annika.gruetzner-zahn@dfki.de, georg.rehm@dfki.de

Maria Giagkou · Stelios Piperidis
R. C. "Athena", Greece, mgiagkou@athenarc.gr, spip@athenarc.gr

[1] This chapter is based on Gaspari et al. (2021, 2022a,b), Giagkou et al. (2022), and Grützner-Zahn and Rehm (2022).

the EU. Against this background, the EU-funded project European Language Equality (ELE) has addressed the issue of Digital Language Equality (DLE) in Europe, with the intention of tackling the imbalances across Europe's languages, that have widened even further in the meantime, as explained in Chapter 4. ELE's contribution to advancing DLE in Europe hinges on a systematically developed and inclusive all-encompassing strategic research, innovation and implementation agenda (SRIA) and a related roadmap to drive forward much needed efforts in this direction (see Chapter 45). The present chapter describes the notion of DLE and the associated metric that are at the heart of these plans, and presents the DLE dashboard that visualises the digital support of each European language, so as to monitor the overall progress towards DLE in Europe, also in a comparative fashion across languages.

Despite the persisting imbalances, Europe has come a long way in recognising and promoting languages as fundamental rights of its people and essential components of its unique combined cultural heritage, and this awareness is reflected in research and policy advancements of the last two decades. Krauwer (2003) represented one of the earliest calls for action towards the development of Language Resources and Technologies (LRTs), in particular for under-resourced languages. In the following years, several projects and initiatives contributed to the progress of Europe's languages in terms of technological and digital support; some of the main efforts in this area that laid the foundation for subsequent substantial progress were, e. g., Euromatrix (Eisele et al. 2008), iTranslate4.eu (Yvon and Hansen 2010), FLaReNet (Soria et al. 2012) and CLARIN (Hinrichs and Krauwer 2014). Additionally, META-NET, an EU Network of Excellence forging the Multilingual Europe Technology Alliance, was established and a group of projects (T4ME, CESAR, METANET4U, META-NORD) promoted and supported the development of Language Technologies (LTs) for all European languages (Rehm and Uszkoreit 2012, 2013; Rehm et al. 2016). The EU project CRACKER (Cracking the Language Barrier, 2015-2017) continued the work of META-NET, concentrating on additional strategy development and community building (Rehm et al. 2020). The most recent EU-funded projects continuing efforts in this area were European Language Grid (ELG, Rehm 2023b) and European Language Equality (ELE, Rehm et al. 2022), which collaborated closely, leading to the development of the DLE Metric and the DLE dashboard presented in this chapter.

2 Related Work

While our work on DLE focused specifically on the languages of Europe, it is located in a broader context of related recent efforts with a wider remit, which are briefly reviewed here to pinpoint issues of interest for the subsequent presentation of the definition of DLE, its metric and the dashboard. Joshi et al. (2020) investigate the relation between the languages of the world and the resources available for them as well as their coverage in Natural Language Processing (NLP) conferences, providing evidence for the severe disparity that exists across languages in terms of technological support and attention paid by academic, scientific and corporate play-

ers. In a similar vein, Blasi et al. (2022, p. 5486) argue that the substantial progress brought about by the generally improved performance of NLP methods "has been restricted to a minuscule subset of the world's approx. 6,500 languages", and present a framework for gauging the global utility of LTs in relation to demand, based on the analysis of a sample of over 60,000 papers published at major NLP conferences. This study also shows convincing evidence for the striking inequality in the development of LTs across the world's languages. While this severe disparity is partly in favour of a few, mostly European, languages, on the whole, the vast majority of the languages spoken in Europe are at a disadvantage.

Simons et al. (2022) develop an automated method to evaluate the level of technological support for languages across the world. Scraping the names of the supported languages from the websites of over 140 tools selected to represent a good level of technological support, they propose an explainable model for quantifying and monitoring digital language support on a global scale. Khanuja et al. (2022) propose an approach to evaluate NLP technologies across the three dimensions of inclusivity, equity and accessibility as a way to quantify the diversity of the users they can serve, with a particular focus on equity as a largely neglected issue. Their proposal consists of addressing existing gaps in LRT provision in relation to societal wealth inequality. Khanuja et al. (2022) lament in particular the very limited diversity of current NLP systems for Indian languages, and to remedy this unsatisfactory situation they demonstrate the value of region-specific choices when building models and creating datasets, also proposing an innovative approach to optimise resource allocation for fine-tuning. They also discuss the steps that can be taken to reduce the biases in LRTs for Indian languages and call upon the community to consider their evaluation paradigm in the interest of enriching the linguistic diversity of NLP applications.

Acknowledging that LTs are becoming increasingly ubiquitous, Faisal et al. (2022) look into the efforts to expand the language coverage of NLP applications. Since a key factor determining the quality of the latest NLP systems is data availability, they study the geographical representativeness of language datasets to assess the extent to which they match the needs of the members of the respective language communities, with a thorough analysis of the striking inequalities. Bromham et al. (2021) examine the effects of a range of demographic and socio-economic aspects on the use and status of the languages of the world, and conclude that language diversity is under threat across the globe, including in industrialised and economically advanced regions. This study finds that half of the languages under investigation faced serious risks of extinction, potentially within a generation, if not imminently. This is certainly an extremely sombre situation to face up to, which calls for a large-scale mobilisation of all possible efforts by all interested parties to avoid such a daunting prospect, particularly in Europe, where multilingualism is recognised as an important part of diversity. Establishing a working definition of DLE, devising a metric to measure the situation of each European language with respect to DLE and implementing an interactive dashboard to monitor progress in this direction are vital elements of this large-scale endeavour.

3 Digital Language Equality: Key Principles and Definition

The DLE Metric and the DLE dashboard can be used to measure, visualise and compare the position of Europe's languages with respect to DLE on the basis of up-to-date and carefully chosen quantitative indicators. In this context, language *equality* does not mean *sameness* on all counts, regardless of the respective environments of the languages; in fact, the different historical developments and current situations of the very diverse languages under consideration are duly taken into account, along with their specific features, different needs and realities of their communities, e. g., in terms of number of speakers, ranges of use, etc., which vary significantly. It would be naive and unrealistic in practice to disregard these facts, and to set out to erase the differences that exist between languages, which are vital reflections of the relevant communities of speakers and key components of Europe's shared cultural heritage. This is also a core value of multilingualism in Europe, where all languages are regarded as inherent components of the cultural and social fabric that connects European citizens in their diversity.

In addition, the notion of DLE stays well clear of any judgement of the political, social and cultural status or value of the languages, insofar as they collectively contribute to a multilingual Europe that should be supported and promoted. Alongside the fundamental concept of *equality*, we also recognise the importance of the notion of *equity*, meaning that for some European languages, and for some of their needs, a targeted effort is necessary to advance the cause of equality. For example, the availability of, and access to, certain resources and services (e. g., to revitalise a language, or to promote education through that language) may be very important for some of Europe's languages, but by and large these are not pressing issues, for instance, for most official national languages. With this in mind, the definition of DLE and the implementation of the DLE Metric discussed below are intended to accurately capture the needs and expectations of the various European languages, and especially the shortfalls with respect to being adequately served in terms of resources, tools and technological services in the digital age, so as to support the large-scale efforts to achieve DLE, also through data analytics and visualisation in the DLE dashboard.

The definition of DLE drew inspiration, among others, from the META-NET White Paper Series (Rehm and Uszkoreit 2012) and from the BLARK concept (Basic Language Resource Kit, Krauwer 2003), which have been instrumental in assessing the level of technological support for specific languages, and in particular in identifying those that lag behind in the digital age and in encouraging the targeted interventions required to fill the gaps in LT support. These starting points were further elaborated by the ELE consortium in collaboration with its vast networks of contacts and partnerships, also in light of the latest developments in LRTs and in language-centric AI techniques and of the evolution of the relevant institutional, academic, industrial and business landscape that has grown and diversified considerably in the last two decades, as discussed in other chapters of this book. Following a systematic and inclusive consultation effort in the ELE consortium, the following consensus was achieved (Gaspari et al. 2021, p. 4).

Digital Language Equality (DLE) is the state of affairs in which all languages have the technological support and situational context necessary for them to continue to exist and to prosper as living languages in the digital age.

This definition was applied to 89 European languages in the project: all 24 official EU languages, 11 additional official national languages and 54 RMLs. This definition, in turn, provided the conceptual basis to design and implement a metric to enable the quantification of the level of technological support of each European language with descriptive, diagnostic and predictive value to promote DLE in practice. This approach allows for comparisons across languages, tracking their progress towards the ultimate collective goal of DLE in Europe, as well as the prioritisation of interventions to meet any needs, especially to fill identified gaps, focusing on realistic and feasible targets, as part of the implementation of the all-encompassing SRIA and related roadmap devised by ELE to drive the advancement towards DLE, as described in detail in Chapter 45.

4 Implementing the Digital Language Equality Metric

Based on the definition of DLE, we describe the associated metric as follows (Gaspari et al. 2021, p. 4):

The Digital Language Equality (DLE) Metric is a measure that reflects the digital readiness of a language and its contribution to the state of technology-enabled multilingualism, tracking its progress towards the goal of DLE.

The DLE Metric is computed for each European language on the basis of a range of quantifiers, grouped into technological factors (TFs, that correspond to the available resources, tools and services, Gaspari et al. 2022a) and situational contextual factors (CFs, that reflect the broad socio-economic ecosystem of each language, which determines the potential for technology and resource development, Grützner-Zahn and Rehm 2022).

The setup and formulation of the metric are modular and flexible, i. e., they consist of well-defined separate and independent, but tightly integrated quantifiers. In particular, the TFs were devised so as to be compatible with the metadata schema adopted by the European Language Grid cloud platform[2] (Labropoulou et al. 2020; Piperidis et al. 2023). The ELG cloud platform bundles together datasets, corpora, functional software, repositories and applications to benefit European society, industry and academia and administration, and provides a convenient single access point to LRTs for Europe's languages (Rehm 2023a).

[2] https://www.european-language-grid.eu

In addition, the definition of DLE and its associated metric have been designed to be transparent and intuitive for linguists, LT experts and developers, language activists, advocates of language rights, industrial players, policy-makers and European citizens at large, to encourage the widest possible uptake and buy-in to the cause of DLE across Europe. In establishing the DLE definition and its associated metric, an effort was made for them to be founded on solid, widely agreed principles, but also striking a balance between a methodologically sound and theoretically convincing approach, and a transparent formulation. The rationale behind this approach was that the DLE definition and its metric should be easily understood and able to inform future language and LT-related policies at the local, regional, national and European levels in order to guide and prioritise future efforts in the creation, development and improvement of LRTs according to the SRIA and roadmap (see Chapter 45), with the ultimate goal of achieving DLE in Europe by 2030.

Through data analytics and visualisation methods in the DLE dashboard (see Section 7), European languages facing similar challenges in terms of LT provision can be grouped together, and requirements can be formulated to support them in remedying the existing gaps and advancing towards full DLE. A crucial feature of the DLE Metric is its dynamic nature, i. e., the fact that its scores can be updated and monitored over time, at regular intervals or whenever one wishes to check the progress or the status of one or more European languages. This is why the DLE Metric is a valuable tool to achieve DLE for all European languages, and a key element of the sustainable evidence-based SRIA and of the roadmap guiding future interventions promoting LTs and language-centric AI across Europe.

5 Technological Factors

In order to objectively quantify the level of technological support for each of Europe's languages, a number of TFs were considered. The following description presents their main categories, illustrating the breadth and diversity of the LRTs that they capture through the ELG catalogue (Rehm 2023a; Piperidis et al. 2023; Labropoulou et al. 2020). In that regard, we assume that the ELG catalogue, with its more than 13,000 LRTs at the time of writing, provides a representative picture of the state of play of technology support of Europe's languages.

The first category of TFs is based on the availability of LRs, i. e., corpora, datasets or collections of text documents, text segments, audio transcripts, audio and video recordings, etc., monolingual or bi-/multilingual, raw or annotated. This category also encompasses language models and computational grammars and resources organised on the basis of lexical or conceptual entries (lexical items, terms, concepts, etc.) with their supplementary information (e. g., grammatical, semantic, statistical information, etc.), such as lexica, gazetteers, ontologies, term lists, thesauri, etc.

The resulting technological DLE score for each European language is a reflection of the LRTs available in the ELG catalogue for that language. While the number of available LRs is an essential aspect of a language's digital readiness, the specific

types and features of these LRs are equally important, insofar as they indicate how well a language is supported in the different LT areas. To capture such aspects in the DLE Metric, in addition to raw counts of available LRs, the following LR features have also been taken into account and attributed specific weights in the scoring mechanism (see Table 1, p. 66, in the Appendix):

- resource type
- resource subclass
- linguality type
- media type covered or supported
- annotation type (where relevant)
- domain covered (where relevant)
- conditions of use

The second category of TFs is based on the availability of tools and services offered via the web or running in the cloud, but also downloadable tools, source code, etc. This category encompasses, for example, NLP tools (morphological analysers, part-of-speech taggers, lemmatisers, parsers, etc.); authoring tools (e. g. spelling, grammar and style checkers); services for information retrieval, extraction, and mining, text and speech analytics, machine translation, natural language understanding and generation, speech technologies, conversational systems, etc. The features of tools and services that are considered and assigned weights in the scoring system of the DLE Metric (see Table 2, p. 67), are as follows:

- language (in)dependent
- type of input processed
- type of output provided
- type of function
- domain covered (where relevant)
- conditions of use

5.1 Weights and Scores

The weights given to the feature values of the LRTs quantify their contribution to the DLE score with regard to the relevant TFs. The scoring system (see Tables 1 and 2) is based on the assumption that for any language some features of LRTs contribute more effectively to achieving DLE than others. Higher weights are assigned to feature values related to 1. more complex LRTs, e. g., tools that process or support more than one modality, 2. more expensive and labour-intensive datasets or tools, e. g., in terms of the effort required to build them, 3. more open or freely available datasets and tools, and 4. additional envisaged applications that could be supported.

One guiding consideration in developing the DLE Metric, and especially in assigning the weights of the features and their values for the TFs, is to make the fewest possible assumptions about the (preferred or supposedly ideal) use-cases and actual

application scenarios that may be most relevant to users. These can vary widely for all languages on the basis of a number of factors impossible to establish a priori. We therefore refrained from predetermining particular preferred end-uses when implementing the full specification of the DLE Metric, which otherwise would risk it being unsuitable for some end-users and applications. Here we briefly review some of the key features of the TFs, focusing on those that can have several values.

For instance, a feature of LRs that can receive several values is that of *Annotation Type*, where applicable. In the implementation of the DLE Metric, we assign a constant very small fixed weight, also based on the fact that some LRs can possess several annotation types in combination. A similar consideration applies to the *Domain* feature (again, where relevant), which has many possible values both for LRs and for tools and services: in these cases, the weights assigned to *Domain* values are fixed and relatively small, again considering that multiple domains can be combined in a single LR, tool or service. In addition to *Domain*, another feature that appears both in LRs and tools and services is *Conditions of use*: the weights proposed for this feature of the TFs are identical for the corresponding values of *Conditions of use* across datasets and tools and services. In the case of (much) more restrictive licensing terms, lower weights are assigned than to liberal use conditions, so they contribute (much) less to the partial technological DLE score for the LRT in question, and therefore to the overall technological DLE score for the specific language.

5.2 Configuration of the Technological Factors

Before coming up with the final implementation of the weighting and scoring system for the TFs (see Tables 1 and 2), we experimented with a range of different setups. We used the contents of the ELG catalogue as of early 2022, which at that time contained about 11,500 records, out of which about 75% were datasets and resources (corpora, lexical resources, models, grammars) and the rest were tools and services. These records contained multiple levels of metadata granularity. The ELG repository had been populated with LRTs following extensive efforts by a wide range of language experts and reflected the input of this community of experts, mobilised in ELE, to ensure comprehensive coverage, which is why we considered the ELG catalogue representative with regard to the existence of LRTs for Europe's languages, so it was used as the empirical basis for the computation of the technological DLE scores.

The ELG catalogue includes metadata for LRs and LTs. In ELG, each resource and tool/service has several features and associated values, based on the schemes presented in Tables 1 and 2. Each feature was initially assigned a tentative weight to calculate preliminary technological DLE scores of each language, comparing the resulting scores of a number of alternative preliminary setups. During this fine-tuning of the weights, we considered especially where each language stood in relation to the others and how their relative positioning changed as a result of assigning different weights to the various feature values. This was an efficient and effective method to

gradually refine the setup of the TFs and propose the implementation of the weights in the scoring mechanism that was eventually adopted (see Tables 1 and 2).

The experiments showed that the global picture of the technological DLE scores for the languages of Europe tended not to change dramatically as the weights assigned to the feature values were manipulated. We experimented both with very moderate and narrow ranges of weights, and with more extreme and differentiated weighting schemes. Since, ultimately, any changes were applied across the board to all LRTs included in the ELG catalogue for all languages, any resulting changes propagated proportionally to the entire set of languages, thus making any dramatic changes rather unlikely, unless one deliberately rewarded (i. e., gamed) features known to disproportionately affect one or more particular languages. It is clear that this would have been a biased and unfair manipulation of the DLE Metric, and was therefore avoided, as we wanted the relevant scores to be a fair, and bias-free, representation of the status of all European languages with respect to DLE.

These preliminary experiments carried out in early 2022 to finalise the setup of the TFs for the DLE Metric demonstrated that the overall distribution of the languages tended to be relatively stable. This was due partly to the sheer amount of features and possible feature values that make up the TFs. As a result, even if one changed the weights, with the exception of minor and local fluctuations, three main phenomena were generally observed while testing the DLE Metric and its TF scores.

1. The overall positioning of the languages remained largely stable, with a handful of languages standing out with the highest technological DLE scores (English leading by far, typically over German, Spanish and French, with the second language having roughly half the technological DLE score of English), the many minimally supported languages still displaying extremely low technological DLE scores, and a large group of similarly supported languages in the middle.
2. Clusters of languages with similar LT support according to intuition and expert opinion remained ranked closely together, regardless of the adjustments made to specific weights for individual features and their values.
3. Even when two similarly supported languages changed relative positions (i. e., language A overtook language B in terms of technological DLE score) as a result of adjusting the weights assigned to specific features and their values, their absolute technological DLE scores still remained very close, and the changes in ranking tended not to affect other neighbouring languages on either side in a noticeable manner.

During the preliminary testing that eventually led to the final setup of the TFs in the DLE Metric presented in Tables 1 and 2, we performed focused checks on pairs or small sets of languages spoken by comparable communities and used in nearby areas or similar circumstances, and whose relative status in terms of LT support is well known to the experts. These focused checks involved, e. g., Basque and Galician, Irish with respect to Welsh, and the dozen local languages of Italy (also with respect to Italian itself), etc. Overall, the general stability and consistency demonstrated by the technological DLE scores across different setups of weight assignments for the various features and their possible values for TFs provided evidence of its validity

as an effective tool to guide developments and track progress towards full DLE for all of Europe's languages. In essence, the setup eventually selected (Tables 1 and 2) ensures that the DLE Metric optimally captures the real situation of all of Europe's languages in the digital age, tracking the progress towards DLE.

5.3 Computing the Technological Scores

Based on the above, the steps to calculate the technological DLE score which is part of the DLE Metric are as follows:

1. Each LRT in the ELG catalogue obtains a score ($Score_{LRT}$), which is equal to the sum of the weights of its relevant features (see Tables 1 and 2 for the weights and associated values). Specifically for features *Annotation Type* and *Domain*, instead of simply adding the respective weight, the weight is multiplied by the number of unique feature values the LR in question has (see Section 5.1).
   Example: Suppose an LRT in the ELG catalogue (LRT1) has the following features: corpus, annotated, monolingual, with three different annotation types (morphology, syntax, semantics), with text as media type, covering one domain (e. g., finance), with condition of use *research use allowed*. Then, using the weights as specified in Table 1, LRT1 is assigned the following score:

$$Score_{LRT1} = 5 + 1 + 2.5 + (3 * 0.25) + 1 + (1 * 0.3) + 3.5 = 14.05$$

2. To compute the technological DLE score for language X ($TechDLE_{LangX}$) we sum up the $Score_{LRT}$ of all LRTs that support language X (LRT1, LRT2, …LRTN), i. e.,

$$TechDLE_{LangX} = \sum_{i=1}^{N} Score_{LRTi}$$

Similarly, any tool or service included in the ELG catalogue receives a partial score with the same procedure, on the basis of the weights presented in Table 2. As the ELG catalogue organically grows over time, the resulting technological DLE scores are constantly updated for all European languages. These scores can be visualised through the DLE dashboard (see Section 7), providing an up-to-date and consistent (i. e., comparable) measurement of the level of LT support and provision that each language of Europe has available, also showing where the status is not ideal or not at the level one might expect.

5.4 Technological DLE Scores of Europe's Languages

Figure 1 shows the technological DLE scores for all of Europe's languages as of late February 2023, obtained on the basis of the final weighting and scoring mechanism described in the previous sections.

Not surprisingly, based on the TFs of the DLE Metric, at the time of writing in early 2023, English is still by far the most well-resourced language of Europe, leading the way over German and Spanish, that follow with very similar technological DLE scores, which are roughly half that of English. French has a marginally lower score, which places it in fourth position. Italian, Finnish and Portuguese follow at some distance, and it is interesting to note that the next cluster of languages that are spoken by sizeable communities in Europe (e. g., Polish, Dutch, Swedish), still in the top ten of the overall list of languages, have a technological DLE score that is roughly six times lower than that of English: a stark reminder based on evidence provided by the ELG catalogue and measured through the DLE Metric of the persisting imbalances in the overall digital support of Europe's languages, showing that urgent decisive action is needed to achieve DLE (Chapter 4 provides a more detailed cross-language comparison).

5.5 Open Issues and Challenges

The technological DLE scores based on the TFs do not take into account the size of the LRs or the quality of the LRTs included in ELG. While these are important features, there exist a large variety of size units for LRs, and the way of measuring data size is not standardised, especially for new types of LRs such as language models. Regarding the quality of tools and services in particular, while some information on the Technology Readiness Level[3] scale is available in ELG, the large number of null values does not make it easy to take this aspect into account for consistency reasons. These are shortcomings that can be revisited in subsequent efforts, with a view to overcoming these limitations and further improving the overall accuracy and granularity of the technological DLE scores going forward.

As far as datasets are concerned, in particular, there could be benefits in setting a minimum size criterion to include LRs such as corpora or grammars in the computation of the technological DLE score, e. g., to avoid using very small resources that cannot be realistically applied in actual technology development scenarios. However, it is difficult to establish arbitrarily what this minimum size threshold should be, also in recognition of the specifics of the languages of Europe. As a result, the decision was made not to set any minimum size requirement for LRs. The thinking behind this choice was that relatively small datasets are common in less-resourced languages, for particular domains, etc., and there is the possibility to merge small datasets to create bigger ones that would, in fact, be useful, for instance in domain

[3] https://en.wikipedia.org/wiki/Technology_readiness_level

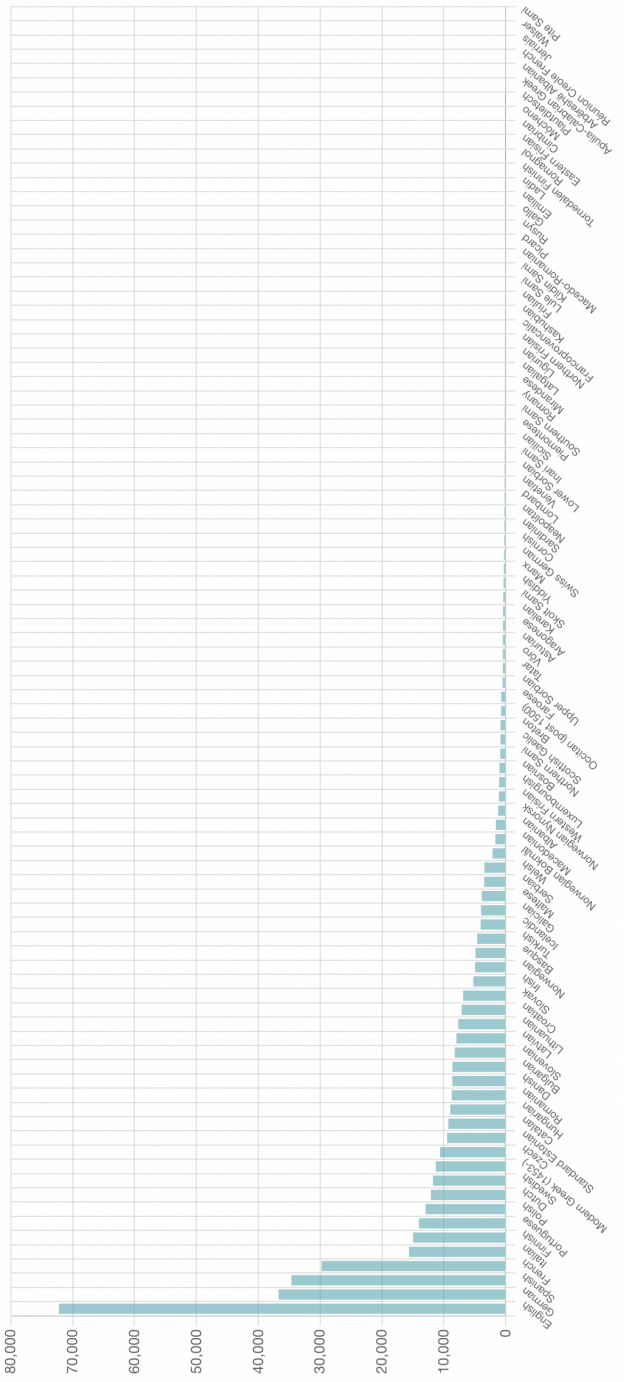

Fig. 1 Technological Digital Language Equality scores as of late February 2023

adaptation for MT, to mention but one example. More broadly, by proposing the DLE Metric we intend to foster a culture of valuing all and any LRTs, especially for less-resourced languages, judiciously balancing the importance given to the size, quantity, diversity and quality of the LRTs, being mindful that several of Europe's languages are in dire need of support.

6 Contextual Factors

While the technological scores based on the TFs represent the technological support of a language, they do not reflect the overall socio-political environment of a language. There are other factors that influence how a language thrives in the digital age, such as political will, funding, being the object of research projects, economic interest, etc. The importance of creating a picture that reflects this environment of a language community was recently also considered by other researchers. Several data-driven studies analyse the relationship between the technical support of a language and non-technological factors (see Section 2).

Related approaches attempt to measure the influence of non-technological factors on the development of LRTs considering often only individual factors in the realm of economy (usually the Gross Domestic Product, GDP), research (e. g., number of publications in specific conferences) and the size of the language community. In the DLE Metric, the Contextual Factors (CFs) are defined as the "general conditions and situations of the broader context" of a language community (Gaspari et al. 2021, p. 7). This definition includes factors from all areas of life assuming that those have an influence on the development and use of LRTs.

Economy Factors in this area reflect the general and the LRT-specific part of the economy. The overall welfare of the language community and the size of the potential market are important factors for companies to invest in the development of LRTs for a language.

Education The language and digital literacy level of a language community influences the use of a language online and on digital devices. Additionally, to be able to develop LRTs, researchers with technical but also linguistic skills of the respective languages are needed.

Funding Investment in research and innovation in the area of LT is necessary for basic and applied research on which technology development is based.

Industry Companies, both well-established and startups, are important drivers of the development and distribution of LT applications, tools and services.

Law The legal framework can hinder progress or steer developments in certain directions.

Media The creation and distribution of news, newspapers, magazines, films, etc. in a language constitutes, on the one hand, a possible large dataset for the development of LRTs, and on the other hand, demonstrates the willingness to make content accessible to the language community.

Online The online representation of a language community indicates that active community members are willing and determined to use the language in the digital world. Additionally, the availability of online data in the respective language gives researchers or developers the opportunity to create LRs.

Policy Strategic plans and agendas at local, regional and national levels indicate the political will to support a topic and the direction in which policy-makers intend to lead society in the future.

Public Administration Public authorities represent the state to its citizens. The inclusion and support of languages spoken in the country or region by public authorities enables participation and utilisation within the society.

Research & Development & Innovation Innovations depend on basic and applied research and on the development of products that are ready for the market. This requires a minimum of research positions in relevant institutions and supporting infrastructure.

Society The social attitude towards a language has a great influence on how much investment, effort and time are put into the preservation of a language by the language community and by the state.

Technology The technological infrastructure reflects the possibility for a language community to access and take a part in the digital world.

6.1 Computing the Contextual Scores

6.1.1 Data Sources and Collection

Initially, 72 potential contextual factors were identified through the collection of factors considered relevant in publications such as, among others, the STOA study (STOA 2018), the META-NET White Paper Series (Rehm and Uszkoreit 2012) and EFNIL's European Language Monitor (ELM);[4] we also consulted with the 52 ELE project partners. The 72 tentative CFs were clustered into 12 areas (see above) representing different aspects of a language's context (Gaspari et al. 2021).

To be measurable, each factor had to be quantified with an indicator, which depended on the existence and accessibility of corresponding data. First, different data sources were collected including, among others, EUROSTAT,[5] ELM, Ethnologue[6] and various reports and articles. Second, possible indicators for each factor were considered and matched with the available data. GDP, for example, was considered to be a suitable indicator for the factor "economic size".

Eventually, 27 of the 72 initial factors had to be excluded due to missing data. This affected especially factors from the areas "research & development & innovation", "society" and "policy". Data about policies is essentially too broad and reflects rather

[4] http://www.efnil.org/projects/elm

[5] https://ec.europa.eu/eurostat

[6] https://www.ethnologue.com

coarsely whether policies exist or not. For instance, the factor "presence of local, regional or national strategic plans, agendas, committees working on the language, LT, NLP, etc." was quantified on data indicating whether a national agenda with regard to AI and LTs exists. Considering also local and regional plans and the existence and maybe also number and size of committees would require much more detailed data. The factors excluded from the class "research & development & innovation" covered mainly figures about the LT research environment, while broader numbers about the research situation of the whole country were indeed available. Tables 4-15 in the Appendix show all factors from the preliminary definition (Gaspari et al. 2021, 2022b), their class and the indicator they were quantified with. Overall, 46 factors were quantified with at least one appropriate indicator, and some with two indicators representing different perspectives like total numbers and numbers per capita.

The data was collected in late 2021. Many sources provided their data as spreadsheets, while some data was published as HTML documents. The data for 15 indicators had to be collected manually from reports and articles. We attempt to update the contextual factors on an annual basis. Preliminary tests indicate that updating the contextual DLE scores for all EU languages takes up to two weeks of work by one member of staff who is familiar with the structure and nature of the CFs.

6.1.2 Data Processing

The collected CF data was very heterogeneous: it had different formats, was based on country or language community level, included differing languages or countries and consisted of different data types. Data preparation took several steps, including data format standardisation, harmonising language names based on Glottolog (Hammarström et al. 2021) and data merging. Some sources provided plain text from which a score had to be manually determined. Features mentioned in the text, e. g., regarding the existence of a national LT policy, were quantified with a number and this number was assigned to countries or language communities. If the text included more than one feature, the numbers were added up, e. g., if a country published several policies covering the topic AI and LTs. Table 3 (p. 68) shows a list of the indicators transformed from plain text.

The DLE Metric processes data on a per-language basis. Thus, data collected on the *country* level had to be converted to the *language* level. In total, the factors were quantified with three different types of data, namely absolute numbers, proportional numbers, and scores. Total numbers were split proportionally, using the percentage of speakers of the language per country. The percentages were calculated through population size and number of speakers. Due to some gaps and old records, experts from the ELE consortium were asked to provide missing or more up-to-date and reliable data. The figures for Alsatian, Faroese, Gallo, Icelandic, Macedonian and the Saami languages were corrected accordingly.

Languages often taught as a second language (English, German, French, Spanish) were only included in the mapping if the language had an official status in the country. For example, the figures for English consist of the figures of the UK, Ireland and

Malta (in other European countries, English does not have official status). If the language was an official national language in at least one country, only language communities with more than one percent were included to simplify the mapping. Total numbers per capita of a language community, proportional numbers, and scores were applied to the language communities without adjustment.

If a language was spoken in more than one country, total numbers were added up, while proportional numbers, scores and total numbers per capita were calculated through the average; the different sizes of the language communities were partly taken into account, hence, the data values of bigger language communities were weighted double for the calculation of the average. However, a more complex inclusion of the size of the language community would result in more fine-grained figures, which would probably affect the contextual DLE scores to some extent.

6.1.3 Calculation of the Contextual Digital Language Equality Score

The data referring to each language community was converted into contextual DLE scores, which indicate the extent to which a language has a context that supports the possibility of evolving digitally or not. Without the political will, funding, innovation and economic interest in the respective region, the probability of achieving DLE is low. Given the underlying complexity, in order for the contextual scores to be easily conceptualised and comparable across languages, a relative score between 0 and 1 was assigned to each language, with 0 representing a context with no potential for the development of LT, and 1 representing the best potential. To keep this part of the DLE Metric as transparent as possible, we decided to base the calculation on an average of the factors. Therefore, the intermediate goal was to calculate a score between 0 and 1 for each factor. The language with the lowest value for the respective factor was attributed 0, while the language with the highest value received 1. The following steps were conducted to calculate the contextual DLE score for each European language:

1. Calculation of the range: highest value – lowest value;
2. $\frac{(value-minimum)*100}{range}$ = Percentage weighting of a language within the range;
3. The result is a relative value: to obtain a score between 0-1 the result is divided by 100;
4. Apply steps 1-3 for all languages and factors;
5. Calculate the average of all factors per language;
6. Weighting of the scores with the three chosen factors of a. number of speakers, b. scores based on the language status, and c. whether the language is an official EU language or not.

The three weighting factors were considered to be particularly relevant for the context to develop LRTs due to the influence of the number of speakers on the investment by large companies and its official status in the EU on the amount of funding. The weighting included two steps: 1. calculating the average of the overall scores, the scores for the number of speakers and the legal status and 2. adding 0.07 to the

score for each official EU language. The second step was separated from the average calculation, because the indicator consisted of two values, 1 if it is an official EU language and 0 if it is not. The average calculation would result in an excessively strong boost for all official EU languages. Hence, with the data for the contextual factors available at the end of 2021, English already had a score of around 0.7-0.8 without the boost. Smaller values for EU languages would have penalised English, which would not have represented reality.

We created five different versions of the possible configurations of the CFs to conduct a thorough comparative evaluation. The factors were classified based on a number of overall properties, i. e., if a data point can be updated automatically or if the data is considered high quality (see Tables 4-15). Data quality was chosen to avoid bias in the overall result caused by extreme maximum and minimum values. For example, for the quantification of the factor "number of podcasts", several platforms were found which could have provided numbers of podcasts in different European languages, but because of different target audiences, the values were highly skewed to the languages spoken by those target audiences. Factors which were quantified with data reflecting no big differences between languages were also excluded by the quality criterion, e. g., the literacy level of all countries varied between 98 and 99 percent, i. e., hardly at all. To be able to update the metric on a regular basis without much manual effort after the end of the ELE project, the possibility of collecting the data fully automatically was picked as the other main criterion.

Based on these criteria, the following CF configurations were examined:

1. Factors with available data: 46 factors
2. Factors that can be updated automatically: 34 factors
3. Factors with good or high data quality: 26 factors
4. Factors that can be updated automatically and that also have good or high data quality: 21 factors
5. A set of manually curated factors using four criteria: automatically updatable, good/high data quality, a maximum of two factors per class, balance between data types: 12 factors (Table 16 shows the factors included in this configuration)

Including fewer factors in the metric increased the risk of omitting an important factor. On the other hand, including fewer factors also reduced the risk of distorting the metric with more data.

6.2 Experts Consultation

Considering that appropriate baselines do not exist, we validated the five different results through the consultation of experts. Individual contextual scores can be interpreted by comparing them to the scores of other languages.

The panel consisted of ELE consortium partners. We selected the members based on their expertise and experience in the areas of LT, Computational Linguistics and Linguistics. Moreover, the experts represented different European countries and

were very familiar with the background of their countries and languages spoken there. We reached out to 37 of the 52 ELE partner organisations. They received the results of the five configurations of the metric and were asked to provide assessments regarding the languages they knew, to explain how they would have expected the results to be, and to indicate the most appropriate configuration.

In total, 18 partners provided assessments. The feedback consisted of overall ratings of the five configurations as well as detailed comments regarding individual languages. As a consequence, most answers related to official EU languages. RMLs for which feedback was received are spoken in the UK, Spain, Italy and the Nordic countries. We received feedback on 56 of the 89 languages.

In general, using all factors was evaluated as risky due to the possible distortion of results caused by data of bad quality. The results of configuration 1 were considered unexpected, with high scores for languages such as Emilian, Gallo and Franco-Provencial, probably caused by distorted data. The second configuration was criticised, too, except for positive comments on the automatic nature of the metric. The results were less distorted but evaluated as worse compared to configurations 3-5. The results of configurations 4 and 5 were similar. Focusing on quality data improved the results significantly. With fewer factors, configuration 5 provided similar results as configuration 4. Configuration 5 was assessed positively regarding the transparency of fewer factors and the possibility to balance the classes.

Overall, the results of the fifth configuration were assessed to represent the context of the language communities in the most adequate way, while there is still room for improvement for a few languages. Table 17 (p. 73) provides more details.

Several suggestions for improvements were made. Since only pan-European data sources were taken into account for reasons of consistency and comparability, one recommendation concerned extending the data through relevant national and regional sources. One expert pointed out that the context of European languages spoken in countries outside of Europe was excluded, and these missing statistics on the development of LRTs would greatly impact the overall scores, e. g., Portuguese in Brazil. Another suggestion referred to missing factors, such as the inclusion of the vitality status of a language being particularly important for RMLs, or the integration of a factor representing competition of a national language with English as the other official national language which often still dominates daily life, e. g., in Ireland, and prevents more widespread use of the other national language in these areas. Another idea was to replace the official EU status as a weighting factor with the country's membership in the European Economic Area (EEA), since these countries also have access to European research funds.

Suggestions were also made regarding the presentation of the results. Language communities having particularly complex political backgrounds are most likely to be misrepresented by a simple calculation based on country-specific data, and should be highlighted and presented with the limits of solely data-driven work for such cases. It was also suggested that languages without a writing system should be emphasised as special cases for the development of LRTs.

Some feedback expressed reservations about the whole approach. A few reviewers pointed out that a single methodology should not be used to take into account

the different complex contexts and realities of Europe's language communities. For example, languages like Maltese, Irish and the other Celtic languages, which scored better than expected according to our experts, are of note here. The relative prosperity of the United Kingdom, even though it is no longer an EU Member State, seems to boost the RMLs spoken in the UK, although in reality these RMLs are strongly dominated by English. The same applies to Ireland, which has a strong economy, a large ICT sector and significant investments in (English) AI and LT research and development, but a very low level of support for Irish LT.

Another point of criticism was the inclusion of data not applied on a per capita basis. As a result, despite having relatively good support, some small language communities were unable to achieve a high score. The size of the language community has an impact on the economic interest, investment, number of researchers, etc. for the language, but for small language communities that have already invested a lot in their language and infrastructure, some of the scores obtained may appear too low compared to the expectations of the experts.

These criticisms can be debated at length, especially in the interest of finding effective solutions to the identified issues, but are very difficult to avoid altogether with such a quantitative approach as the one that is required to define and measure the CFs as part of the DLE Metric.

These first stable results for the CF calculation were improved based on a more fine-grained data mapping from country to language community level and the feedback of the experts. The aggregation of data points from different countries for languages spoken in several countries, e. g., French, was based on the average with a boost for the data points collected from the countries in which the language has an official national status. This process was replaced by the calculation of a weighted average based on the number of speakers of the language communities which reflects the distribution of the language communities better and prevents distortion through too small or too big language communities. In addition, the boost for EU Member States was changed to a boost for countries in the EEA, the vitality status was added as a penalty for declining languages, and those competing with English as the other dominant official national language were also penalised. The results of this adaptation decreased the number of languages that eventually achieved an excessively high contextual score.

6.3 Contextual DLE Scores of Europe's Languages

In all examined configurations, the top third is dominated by the official EU languages, while the RMLs are part of the long tail to the right. Official national languages which are not official EU languages are ranked between the official EU languages and the RMLs. Figure 2 shows the final results after the adaptation.

As expected, English has the best context for the development of LRTs by far. It is followed by German and French. Italian and Spanish are shown in positions 4 and 5. The position of Spanish *after* Italian is caused by the inclusion of data from

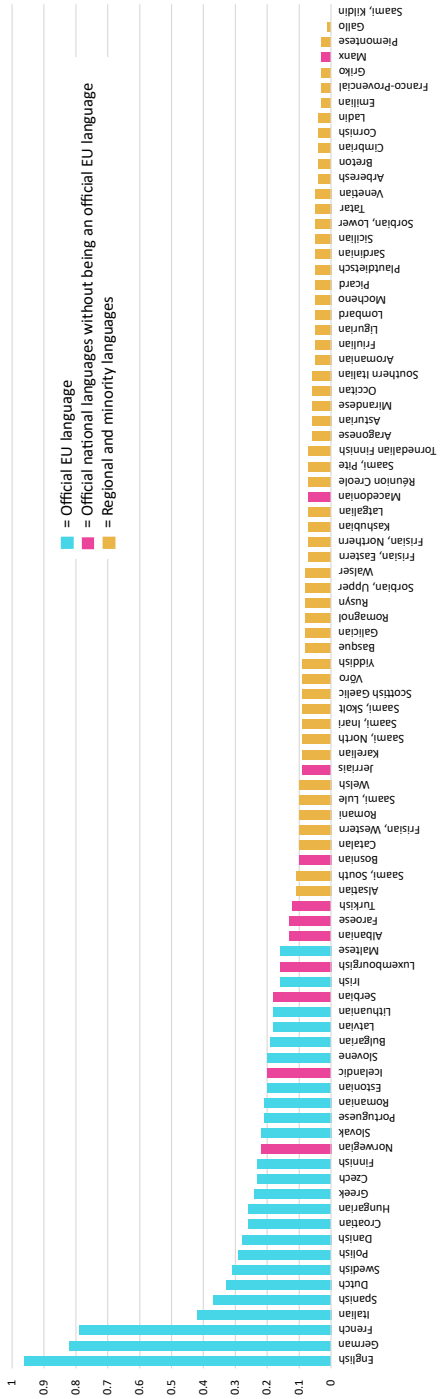

Fig. 2 Contextual Digital Language Equality scores as of late February 2023

European countries only. If data had been included from countries outside of Europe, Spanish, Portuguese, French and English would have had much higher scores. After the five leading languages, variations between the different configurations can be seen. Swedish, Dutch, Danish, Polish, Croatian, Hungarian and Greek are ranked in the upper half of the official EU languages. The official EU languages with the lowest scores are Latvian, Lithuanian, Bulgarian, Romanian, Maltese and Irish which joined this group after the last adjustment.

Among the group of official national languages which are not official EU languages, Norwegian, Icelandic and Serbian are the top performers, achieving contextual DLE scores in line with the middle- and lower-scoring official EU languages, while Manx[7] is presented as a downward outlier. Languages such as Norwegian, Luxembourgish, Faroese and Icelandic achieve better scores than Albanian, Turkish, Macedonian and Bosnian.

The RMLs are led by languages spoken in the more Northern countries like some Saami languages, Western Frisian and Welsh or languages spoken by quite big language communities like Catalan. A total of 23 RMLs achieve contextual DLE scores equal to or lower than 0.05 in the final results, while 30 of the languages obtain scores between 0.06 and 0.1. Kildin Saami and Griko are the languages with the lowest scores.

6.4 Open Issues and Challenges

The contextual DLE scores calculated have some limitations (see Section 6.2). First, expanding the dataset to include regional or national sources would result in 1. a higher number of factors, 2. improved data quality, as the gaps in individual indicators may be filled, 3. quantification of more factors with more than one indicator, to reflect different perspectives, and 4. a more complex mapping to language communities based on regional data resulting in a significant impact on RMLs.

Second, the data cleaning procedure can be improved. One possibility would be to replace outliers with values outside twice the standard deviation by the respective maximum or minimum values of the data series. Data gaps could be filled using data from previous years and skewed data could be corrected using a square root transformation. These processing steps could decrease the impact of distorted data.

An improvement of the mapping from country level to language level could represent regional or urban-rural divides more accurately, especially for larger countries. In particular, the missing mapping of proportional data, scores and total numbers per capita has a major impact on the resulting contextual DLE scores. Here, regional data could help calculate the average deviation of individual regions or language commu-

[7] Manx and Jerriais have been assigned to the group of national languages without being an official EU language, as both languages are recognised as official languages of Jersey and the Isle of Man. Neither island is part of the United Kingdom, but crown dependencies. Therefore, the two languages can be considered both official national languages or RMLs.

nities from other proportional data and to transfer this deviation to proportional data only found on the national level, and similarly for the total figures per capita.

Romaine (2017, p. 49) stresses the importance of an "on-going monitoring of individual communities" for a reliable evaluation of the situation regarding language diversity, which was taken into account with the inclusion of the criterion of automatic updatability of the factors. One problem concerns the eventual interdependencies of the values: the scores of *all* languages may change if new values for some language communities are added, even if the situation of another language community itself has not changed. A temporal dimension could be added to mitigate this.

7 Digital Language Equality Dashboard

In order to provide a precise and easy-to-use tool for presenting and monitoring the TFs and CFs that contribute to the DLE Metric, we designed and implemented a web-based dashboard as part of the European Language Grid.[8] It is available at:

https://live.european-language-grid.eu/catalogue/dashboard

The dashboard shows the contents of the ELG database as interactive visuals dynamically created by user queries, thus providing constantly up-to-date and consistent (i. e., comparable) measurements of the level of LT support and provision across all of Europe's languages (Figure 3). The dashboard provides the figures, statistics and graphs, as appropriate, for:

- the TFs and CFs of the DLE Metric, calculated according to the detailed technical description presented above;
- LRTs hosted in the ELG catalogue, which constitute the source/base data for the TFs that are at the basis of the technological DLE score.

Architecturally, the DLE dashboard consists of two layers: the database of the ELG catalogue and the frontend. The ELG database contents are indexed and saved in JSON. Each user query retrieves the respective results from JSON and exposes them to the front end. While the TFs are calculated dynamically (see Section 5.3) and they reflect the status of the ELG catalogue's database at the time of accessing the dashboard, in the current implementation the CFs are calculated offline, stored in a separate file and exposed to the respective tab of the dashboard's frontend.

[8] https://www.european-language-grid.eu

Fig. 3 DLE dashboard showing the technological (top) and contextual DLE scores (bottom)

8 Conclusions and Future Work

This chapter has introduced the definition of DLE adopted in ELE and has described the DLE Metric, explaining the roles and setups of the complementary TFs and CFs and how the scores are computed. By providing an empirically-grounded and realistic quantification of the level of technological support of the languages of Europe, the DLE Metric is intended to contribute to future efforts to level up the digital support of all of Europe's languages, most notably with the implementation of the evidence-based SRIA and roadmap that will drive future efforts in equipping all European languages with the LRTs needed to achieve full DLE (see Chapter 45). The DLE Metric provides a transparent means to track and monitor the actual progress in this direction, as the technological and contextual DLE scores can be visualised through the DLE dashboard.

The overview of the TFs and CFs is accompanied by discussions of the scoring and weighting mechanisms adopted for the computation of the technological and contextual DLE scores, following extensive testing and expert consultations comparing alternative setups. The chapter explains the overall design of the features and their values with the scores and weighting mechanisms that contribute to the DLE Metric scores, based on data included in the ELG catalogue and the factors eventually selected to represent the specific ecosystems of the languages and their communities. As a result of this, the notion of DLE and its associated metric introduced in this chapter represent valuable tools on which to base future efforts to measure and improve the readiness of Europe's languages for the digital age, also taking into account the situational contexts in which the various languages are used via the CFs.

Thanks to the descriptive, diagnostic and predictive value of the DLE Metric, the community now has a solid and verifiable means of pursuing and evaluating much-needed developments in the interest of all languages of Europe and their speakers. The DLE Metric is relevant to a wide range of stakeholders at local, regional, national and European levels who are committed to preventing the extinction of European languages under threat and who are interested in promoting their prosperity for the future. Such stakeholders include decision- and policy-makers, industry leaders, researchers, developers, and citizens across Europe who will drive forward future developments in the fields of LT and language-centric AI in the interest of DLE.

References

Blasi, Damian, Antonios Anastasopoulos, and Graham Neubig (2022). "Systematic Inequalities in Language Technology Performance across the World's Languages". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, pp. 5486–5505. DOI: 10.18653/v1/2022.acl-long.376. https://aclanthology.org/2022.acl-long.376.

Bromham, Lindell, Russell Dinnage, Hedvig Skirgård, Andrew Ritchie, Marcel Cardillo, Felicity Meakins, Simon Greenhill, and Xia Hua (2021). "Global predictors of language endangerment

and the future of linguistic diversity". In: *Nature Ecology & Evolution* 6, pp. 163–173. https://doi.org/10.1038/s41559-021-01604-y.

Eisele, Andreas, Christian Federmann, Hans Uszkoreit, Herve Saint-Amand, Martin Kay, Michael Jellinghaus, Sabine Hunsicker, Teresa Herrmann, and Yu Chen (2008). "Hybrid machine translation architectures within and beyond the EuroMatrix project". In: *Proceedings of the 12th Annual conference of the European Association for Machine Translation, EAMT 2008, Hamburg, Germany, September 22-23, 2008*. Ed. by John Hutchins, Walther Hahn, and Bente Maegaard. European Association for Machine Translation, pp. 27–34. https://aclanthology.org/2008.eamt-1.6/.

Faisal, Fahim, Yinkai Wang, and Antonios Anastasopoulos (2022). "Dataset Geography: Mapping Language Data to Language Users". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, pp. 3381–3411. DOI: 10.18653/v1/2022.acl-long.239. https://aclanthology.org/2022.acl-long.239.

Gaspari, Federico, Owen Gallagher, Georg Rehm, Maria Giagkou, Stelios Piperidis, Jane Dunne, and Andy Way (2022a). "Introducing the Digital Language Equality Metric: Technological Factors". In: *Proceedings of the Workshop Towards Digital Language Equality (TDLE 2022; co-located with LREC 2022)*. Ed. by Itziar Aldabe, Begoña Altuna, Aritz Farwell, and German Rigau. Marseille, France, pp. 1–12. http://www.lrec-conf.org/proceedings/lrec2022/workshops/TDLE/pdf/2022.tdle-1.1.pdf.

Gaspari, Federico, Annika Grützner-Zahn, Georg Rehm, Owen Gallagher, Maria Giagkou, Stelios Piperidis, and Andy Way (2022b). *Deliverable D1.3 Digital Language Equality (full specification)*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166 ELE. https://european-language-equality.eu/reports/DLE-definition.pdf.

Gaspari, Federico, Andy Way, Jane Dunne, Georg Rehm, Stelios Piperidis, and Maria Giagkou (2021). *Deliverable D1.1 Digital Language Equality (preliminary definition)*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/DLE-preliminary-definition.pdf.

Giagkou, Maria, Penny Labropoulou, Stelios Piperidis, Miltos Deligiannis, Athanasia Kolovou, and Leon Voukoutis (2022). *Deliverable D1.37 Database and Dashboard*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/DLE-dashboard.pdf.

Grützner-Zahn, Annika and Georg Rehm (2022). "Introducing the Digital Language Equality Metric: Contextual Factors". In: *Proceedings of the Workshop Towards Digital Language Equality (TDLE 2022; co-located with LREC 2022)*. Ed. by Itziar Aldabe, Begoña Altuna, Aritz Farwell, and German Rigau. Marseille, France, pp. 13–26. http://www.lrec-conf.org/proceedings/lrec2022/workshops/TDLE/pdf/2022.tdle-1.2.pdf.

Hammarström, Harald, Robert Forkel, Martin Haspelmath, and Sebastian Bank (2021). *Glottolog 4.5*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.5772642.

Hinrichs, Erhard and Steven Krauwer (2014). "The CLARIN Research Infrastructure: Resources and Tools for eHumanities Scholars". In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland: European Language Resources Association (ELRA), pp. 1525–1531. http://www.lrec-conf.org/proceedings/lrec2014/pdf/415_Paper.pdf.

Joshi, Pratik, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury (2020). "The State and Fate of Linguistic Diversity and Inclusion in the NLP World". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*. Online: Association for Computational Linguistics, pp. 6282–6293. DOI: 10.18653/v1/2020.acl-main.560. https://aclanthology.org/2020.acl-main.560.

Khanuja, Simran, Sebastian Ruder, and Partha Talukdar (2022). *Evaluating Inclusivity, Equity, and Accessibility of NLP Technology: A Case Study for Indian Languages*. DOI: 10.48550/ARXIV.2205.12676. https://arxiv.org/abs/2205.12676.

Krauwer, Steven (2003). "The Basic Language Resource Kit (BLARK) as the First Milestone for the Language Resources Roadmap". In: *Proceedings of the International Workshop Speech and Computer (SPECOM 2003)*. Moscow, Russia.

Labropoulou, Penny, Katerina Gkirtzou, Maria Gavriilidou, Miltos Deligiannis, Dimitris Galanis, Stelios Piperidis, Georg Rehm, Maria Berger, Valérie Mapelli, Michael Rigault, Victoria Arranz, Khalid Choukri, Gerhard Backfried, José Manuel Gómez Pérez, and Andres Garcia-Silva (2020). "Making Metadata Fit for Next Generation Language Technology Platforms: The Metadata Schema of the European Language Grid". In: *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*. Ed. by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. Marseille, France: ELRA, pp. 3421–3430. https://www.aclweb.org/anthology/2020.lrec-1.420/.

Piperidis, Stelios, Penny Labropoulou, Dimitris Galanis, Miltos Deligiannis, and Georg Rehm (2023). "The European Language Grid Platform: Basic Concepts". In: *European Language Grid: A Language Technology Platform for Multilingual Europe*. Ed. by Georg Rehm. Cham: Springer, pp. 13–36. DOI: 10.1007/978-3-031-17258-8_2. https://doi.org/10.1007/978-3-031-17258-8_2.

Rehm, Georg, ed. (2023a). *European Language Grid: A Language Technology Platform for Multilingual Europe*. Cognitive Technologies. Cham, Switzerland: Springer.

Rehm, Georg (2023b). "European Language Grid: Introduction". In: *European Language Grid: A Language Technology Platform for Multilingual Europe*. Ed. by Georg Rehm. Cognitive Technologies. Cham, Switzerland: Springer, pp. 1–10.

Rehm, Georg, Federico Gaspari, German Rigau, Maria Giagkou, Stelios Piperidis, Annika Grützner-Zahn, Natalia Resende, Jan Hajic, and Andy Way (2022). "The European Language Equality Project: Enabling digital language equality for all European languages by 2030". In: *The Role of National Language Institutions in the Digital Age – Contributions to the EFNIL Conference 2021 in Cavtat*. Ed. by Željko Jozić and Sabine Kirchmeier. Budapest, Hungary: Nyelvtudományi Kutatóközpont, Hungarian Research Centre for Linguistics, pp. 17–47.

Rehm, Georg, Katrin Marheinecke, Stefanie Hegele, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Khalid Choukri, Andrejs Vasiļjevs, Gerhard Backfried, Christoph Prinz, José Manuel Gómez Pérez, Luc Meertens, Paul Lukowicz, Josef van Genabith, Andrea Lösch, Philipp Slusallek, Morten Irgens, Patrick Gatellier, Joachim Köhler, Laure Le Bars, Dimitra Anastasiou, Albina Auksoriūtė, Núria Bel, António Branco, Gerhard Budin, Walter Daelemans, Koenraad De Smedt, Radovan Garabík, Maria Gavriilidou, Dagmar Gromann, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Lindén, Bernardo Magnini, Jan Odijk, Maciej Ogrodniczuk, Eiríkur Rögnvaldsson, Mike Rosner, Bolette Pedersen, Inguna Skadina, Marko Tadić, Dan Tufiş, Tamás Váradi, Kadri Vider, Andy Way, and François Yvon (2020). "The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe". In: *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*. Ed. by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. Marseille, France: ELRA, pp. 3315–3325. https://www.aclweb.org/anthology/2020.lrec-1.407/.

Rehm, Georg and Hans Uszkoreit, eds. (2012). *META-NET White Paper Series: Europe's Languages in the Digital Age*. 32 volumes on 31 European languages. Heidelberg etc.: Springer.

Rehm, Georg and Hans Uszkoreit, eds. (2013). *The META-NET Strategic Research Agenda for Multilingual Europe 2020*. Heidelberg etc.: Springer. http://www.meta-net.eu/vision/reports/meta-net-sra-version_1.0.pdf.

Rehm, Georg, Hans Uszkoreit, Sophia Ananiadou, Núria Bel, Audronė Bielevičienė, Lars Borin, António Branco, Gerhard Budin, Nicoletta Calzolari, Walter Daelemans, Radovan Garabík, Marko Grobelnik, Carmen García-Mateo, Josef van Genabith, Jan Hajič, Inma Hernáez, John Judge, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Lindén, Bernardo Magnini, Joseph Mariani, John McNaught, Maite Melero, Monica Monachini, Asunción Moreno, Jan Odjik, Maciej Ogrodniczuk, Piotr Pęzik, Stelios Piperidis, Adam Przepiórkowski, Eiríkur Rögnvalds-

son, Mike Rosner, Bolette Sandford Pedersen, Inguna Skadiņa, Koenraad De Smedt, Marko Tadić, Paul Thompson, Dan Tufiş, Tamás Váradi, Andrejs Vasiļjevs, Kadri Vider, and Jolanta Zabarskaite (2016). "The Strategic Impact of META-NET on the Regional, National and International Level". In: *Language Resources and Evaluation* 50.2, pp. 351–374. DOI: 10.1007/s10579-015-9333-4. http://link.springer.com/article/10.1007/s10579-015-9333-4.

Romaine, Suzanne (2017). "Language Endangerment and Language Death". In: *The Routledge Handbook of Ecolinguistics*. Abingdon, Oxfordshire: Routledge, pp. 40–55. DOI: 10.4324/9781315687391.ch3. https://www.routledgehandbooks.com/doi/10.4324/9781315687391.ch3.

Simons, Gary F., Abbey L. Thomas, and Chad K. White (2022). *Assessing Digital Language Support on a Global Scale*. DOI: 10.48550/ARXIV.2209.13515. https://arxiv.org/abs/2209.13515.

Soria, Claudia, Núria Bel, Khalid Choukri, Joseph Mariani, Monica Monachini, Jan Odijk, Stelios Piperidis, Valeria Quochi, and Nicoletta Calzolari (2012). "The FLaReNet Strategic Language Resource Agenda". In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, May 23-25, 2012*. Ed. by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis. European Language Resources Association (ELRA), pp. 1379–1386. http://www.lrec-conf.org/proceedings/lrec2012/summaries/777.html.

STOA (2018). *Language equality in the digital age – Towards a Human Language Project*. STOA study (PE 598.621), IP/G/STOA/FWC/2013-001/Lot4/C2. https://data.europa.eu/doi/10.2861/136527.

Yvon, François and Viggo Hansen (2010). "iTranslate4.eu: Internet translators for all European languages". In: *Proceedings of the 14th Annual conference of the European Association for Machine Translation, EAMT 2010, Saint Raphaël, France, May 27-28, 2010*. European Association for Machine Translation. https://aclanthology.org/2010.eamt-1.41/.

Appendix

Feature	Value	Weight
Resource Type	Corpus	5
	Lexical conceptual resource	1.5
	Language description	3.5
Subclass	Raw corpus	0.1
	Annotated corpus	2.5
	Computational lexicon	2
	Morphological lexicon	3
	Terminological resource	3.5
	Wordnet	4
	Framenet	4
	Model	5
	Each of the others (there are 15 more)	0.5
Linguality Type	Multilingual	5
	Bilingual	2
	Monolingual	1
Media Type	Text	1
	Image	3
	Video	5
	Audio	2.5
	Numerical text	1.75
Annotation Type	*Each of these – can be combined in a single LR*	0.25
Domain	*Each of these – can be combined in a single LR*	0.3
Conditions of Use	Other specific restrictions	0.5
	Commercial uses not allowed	1
	No conditions	5
	Derivatives not allowed	1.5
	Redistribution not allowed	2
	Research use allowed	3.5

Table 1 Weights assigned to the technological factors of the DLE Metric for language resources

Feature	Value	Weight
Language Independent	False	5
	True	1
Input Type	Input text	2
	Input audio	5
	Input image	7.5
	Input video	10
	Input numerical text	2.5
Output Type	Output text	2
	Output audio	5
	Output video	10
	Output image	7.5
	Output numerical text	2.5
Function Type	Text processing	3
	Speech processing	10
	Information extraction and information retrieval	7.5
	Translation technologies	12
	Human-computer interaction	15
	Natural language generation	20
	Support operation	1
	Image/video processing	13
	Other	1
	Unspecified	1
Domain	*Each of these – can be combined in a single tool*	0.5
Conditions of Use	Unspecified	0
	Other specific restrictions	0.5
	No conditions	5
	Commercial uses not allowed	1
	Derivatives not allowed	1.5
	Redistribution not allowed	2
	Research use allowed	3.5

Table 2 Weights assigned to the technological factors of the DLE Metric for tools and services

Factor	Merging of the Scores	Conversion from Text to Scores
Public funding available for LTs	Adding up scores for each country	1 for regional funding 1 for national funding 1 for intranational funding 1 for ESIF 1 for EUREKA 1 for EUROSTAT
Legal status and legal protection	Adding up scores per language	10 for statutory national language 10 for de facto national working language 2 for statutory provincial language 2 for statutory provincial working language 1 for recognised language
Publicly available media outcomes	Adding up two scores: one score for language transfer practices for cinema works screened and one for television works broadcast	2 for dub 1.5 for voice over 1.5 for sub and dub 1 for sub
	Adding up scores + division by the number of answers	Broadcast in original language: 5 for mostly/always, 2.5 for sometimes Broadcast with dubbing: 4 for mostly/always, 2 for sometimes Broadcast in original language with voice-over: 3 for mostly/always, 1.5 for sometimes Dual-channel sound: 2 for mostly/always, 1 for sometimes Broadcast with subtitles: 1 for mostly/always, 0.5 for sometimes
Presence of local, regional or national strategic plans	One of the scores per country	1 for no plan/strategy 2 for a plan without mentioning LT 3 for a plan mentioning LT 4 for a plan mentioning LT and minority and regional languages
Political activity	Adding up scores per country	1 score for each document 1 score for each document mentioning LT 2 for each document exclusively about LT 1 for a document covering a specific language 2 for each document published 2020/2021 1 for each document published 2019/2018

Table 3 Contextual factors: Conversion from plain text into scores

ECONOMY	
Factor	**Indicator**
Size of the economy	Annual GDP
	GDP per capita* **
Size of the LT/NLP market	LT market in million Euro
Size of the language service, translating or interpreting market	Number of organisations from the industry in the ELG catalogue* **
Size of the IT/ICT sector	Perc. of the ICT sector in the GDP* **
	ICT service exports in balance of payment* **
Investment instruments into AI/LT	GDE on R&D in relevant areas*
Regional/national LT market	No indicator found
Average socio-economic status	Annual net earnings, 1.0 FTE worker* **
	Life expectancy at age 60**

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 4 Contextual factors: Proposed factors for class "Economy"

EDUCATION	
Factor	**Indicator**
Higher Education Institutions operating in the language	No indicator found
Higher education in the language	No indicator found
Academic positions in relevant areas	Head count of R&D personnel
Academic programmes in relevant areas	No indicator found
Literacy level	Literacy rate*
Students in language/LT/NLP curricula	Total no. of students in relevant areas* **
Equity in education	Proportional tertiary educ. attainment* **
Inclusion in education	Percentage of foreigners attaining tertiary education* **

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 5 Contextual factors: Proposed factors for class "Education"

FUNDING	
Factor	**Indicator**
Funding available for LT research projects	No. of projects funded in relevant areas*
	Score from the national funding programmes
Venture capital available	Venture capital amounts in Euro
Public funding for interoperable platforms	Number of platforms**

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 6 Contextual factors: Proposed factors for class "Funding"

INDUSTRY	
Factor	**Indicator**
Companies developing LTs	No. of enterprises in the ICT area* **
Start-ups per year	Percentage of "Enterprise births"**
Start-ups in LT/AI	Number of AI start ups* **

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 7 Contextual factors: Proposed factors for class "Industry"

LAW	
Factor	**Indicator**
Copyright legislation and regulations	No indicator found
Legal status and legal protection	Scores out of the legal status* **

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 8 Contextual factors: Proposed factors for class "Law"

MEDIA	
Factor	**Indicator**
Subtitled or dubbed visual media	Scores out of language transfer practices* Scores out of answers about broadcast practices
Transcribed podcasts	Number of entries in the CBA*

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 9 Contextual factors: Proposed factors for class "Media"

ONLINE	
Factor	**Indicator**
Digital libraries	Percentage of contribution to Europeana
Impact of language barriers on e-commerce	Percentage of population buying cross-border**
Digital literacy	No indicator found
Wikipedia pages	Number of articles in Wikipedia* **
Websites exclusively in the language	No indicator found
Websites in the language (not exclusively)	Perc. of websites in the languages* **
Web pages	No indicator
Ranking of websites delivering content	12 selected websites supporting the languages
Labels and lemmas in knowledge bases	Number of lexemes in Wikipedia* **
Language support gaps	Language matrix of supported features*
Impact on E-commerce websites	T-Index*

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 10 Contextual factors: Proposed factors for class "Online"

POLICY	
Factor	**Indicator**
Presence of strategic plans, agendas, etc.	Scores out of a list of the published national AI strategies Scores from questionnaire about strategies
Promotion of the LR ecosystem	No indicator found
Consideration of bodies for the LR citation	No indicator found
Promotion of cooperation	No indicator found
Public and community support for resource production best practices	No indicator found
Policies regarding BLARKs	No indicator found
Political activity	Scores out of the list of documents

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 11 Contextual factors: Proposed factors for class "Policy"

PUBLIC ADMINISTRATION	
Factor	**Indicator**
Languages of public institutions	No. of constitutions written in the language
Available public services in the language	Percentage of a maximum score about digital public services** Score for digital public services**

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 12 Contextual factors: Proposed factors for class "Public administration"

RESEARCH & DEVELOPMENT & INNOVATION	
Factor	**Indicator**
Innovation capacity	Innovation Index* **
Research groups in LT	Number of research organisations
Research groups/companies predominantly working on the respective language	No indicator found
Research staff involved in LT	No indicator found
Suitably qualified Research staff in LT	No indicator found
Capacity for talent retention in LT	No indicator found
State of play of NLP/AI	No indicator found
Scientists working in LT/on the language	Number of researchers in relevant areas*
Researchers whose work benefits from LRs and LTs	No indicator found
Overall research support staff	Head count of research support staff* **
Scientific associations or general scientific and technology ecosystem	No indicator found
Papers about LT and or the language	Number of papers about LT** Number of papers about the language* **

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 13 Contextual factors: Proposed factors for class "Research & Development & Innovation"

SOCIETY	
Factor	**Indicator**
Importance of the language	No indicator found
Fully proficient (literate) speakers	Number of L1 speakers*
Digital skills	Perc. of individuals with basic digital skills* **
Size of language community	Total number of speakers* **
Population not speaking the official language(s)	No indicator found
Official or recognized languages	Total no. of languages with official status*
	Number of bordering languages
Community languages	Number of community languages*
Time resources of the language community	No indicator found
Society stakeholders for the language	No indicator found
Speakers' attitudes towards the language	Total number of participants wanting to acquire the language
Involvement of indigenous peoples	No indicator found
Sensitivity to barriers	No indicator found
Usage of social media or networks	Total number of social media users* **
	Percentage of social media users* **

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 14 Contextual factors: Proposed factors for class "Society"

TECHNOLOGY	
Factor	**Indicator**
Open-source technologies of LTs	No indicator found
Access to computer, smartphone etc.	Perc. of households with a computer* **
Digital connectivity and internet access	Perc. of households with broadband* **

*Indicator marked * is automatically updateable – Indicator marked ** provides good quality data*

Table 15 Contextual factors: Proposed factors for class "Technology"

Class	Factor
Economy	Size of economy
	Size of the ICT sector
Education	Students in LT/language
	Inclusion in education
Industry	Companies developing LTs
Law	Legal status and legal protection
Online	Wikipedia pages
R & D & I	Innovation capacity
	Number of papers
Society	Size of language community
	Usage of social media
Technology	Digital connectivity, internet access

Table 16 Contextual factors included in the final configuration (configuration 5)

Appropriate	Ranked too high	Ranked too low	Contrary Opinion
English	Irish	Norwegian	French
Dutch	Italian	Spanish	German
Danish	Swedish	Portuguese	Saami, Northern
Polish	Hungarian	Czech	Latvian
Greek	Croatian	Romanian	
Finnish	Maltese	Bulgarian	
Estonian	Faroese	Icelandic	
Slovene	Scottish Gaelic	Emilian	
Slovak	Cornish	Sicilian	
Lithuanian	Manx		
Serbian	Saami, Southern		
Basque	Saami, Pite		
Catalan	Saami, Lule		
Galician	Saami, Skolt		
Asturian	Saami, Inari		
Aragonese	Sardinian		
Welsh	Romagnol		
Griko			
Lombard			
Ligurian			
Venetian			
Southern Italian			
Friulian			
Piemontese			
Ladin			
25	17	9	4

Table 17 Contextual factors: Assessment of the languages in the final configuration (configuration 5) by the panel of experts