



# EUROPEAN LANGUAGE EQUALITY

**D1.39**

## Report on the West Frisian Language

---

Authors Charlie Robinson-Jones, Ydwine R. Scarse

---

Dissemination level Public

---

Date 30-06-2022

---

## About this document

Project	European Language Equality (ELE)
Grant agreement no.	LC-01641480 – 101018166 ELE
Coordinator	Prof. Dr. Andy Way (DCU)
Co-coordinator	Prof. Dr. Georg Rehm (DFKI)
Start date, duration	01-01-2021, 18 months
Deliverable number	D1.39
Deliverable title	Report on the West Frisian Language
Type	Report
Number of pages	28
Status and version	Final ( <i>Note: this document is not a contractual ELE deliverable.</i> )
Dissemination level	Public
Date of delivery	30-06-2022
Work package	WP1: European Language Equality – Status Quo in 2020/2021
Task	Task 1.3 Language Technology Support of Europe’s Languages in 2020/2021
Authors	Charlie Robinson-Jones, Ydwine R. Scarse
Reviewers	Jane Dunne, Maria Giagkou
Editors	Maria Giagkou, Stelios Piperidis, Georg Rehm, Jane Dunne
EC project officers	Susan Fraser, Miklos Druskoczi
Contact	European Language Equality (ELE) ADAPT Centre, Dublin City University Glasnevin, Dublin 9, Ireland  Prof. Dr. Andy Way – andy.way@adaptcentre.ie  European Language Equality (ELE) DFKI GmbH Alt-Moabit 91c, 10559 Berlin, Germany  Prof. Dr. Georg Rehm – georg.rehm@dfki.de  <a href="http://www.european-language-equality.eu">http://www.european-language-equality.eu</a>  © 2022 ELE Consortium

## Consortium

1	Dublin City University (Coordinator)	DCU	IE
2	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Co-coordinator)	DFKI	DE
3	Univerzita Karlova (Charles University)	CUNI	CZ
4	Athina-Erevnitiko Kentro Kainotomias Stis Plioroforias, Ton Epikoinonion Kai Tis Gnosis	ILSP	GR
5	Universidad Del Pais Vasco/ Euskal Herriko Unibertsitatea (University of the Basque Country)	UPV/EHU	ES
6	CROSSLANG NV	CRSLNG	BE
7	European Federation of National Institutes for Language	EFNIL	LU
8	Réseau européen pour l'égalité des langues (European Language Equality Network)	ELEN	FR
9	European Civil Society Platform for Multilingualism	ECSPM	DK
10	CLARIN ERIC – Common Language Resources and Technology Infrastructure as a European Research Infrastructure Consortium	CLARIN	NL
11	Universiteit Leiden (University of Leiden)	ULEI	NL
12	Eurescom (European Institute for Research and Strategic Studies in Telecommunications GmbH)	ERSCM	DE
13	Stichting LIBER (Association of European Research Libraries)	LIBER	NL
14	Wikimedia Deutschland (Gesellschaft zur Förderung freien Wissens e.V.)	WMD	DE
15	Tilde SIA	TILDE	LV
16	Evaluations and Language Resources Distribution Agency	ELDA	FR
17	Expert System Iberia SL	EXPSYS	ES
18	HENSOLDT Analytics GmbH	HENS	AT
19	Xcelerator Machine Translations Ltd. (KantanMT)	KNTN	IE
20	PANGEANIC-B. I. Europa SLU	PAN	ES
21	Semantic Web Company GmbH	SWC	AT
22	SIRMA AI EAD (Ontotext)	ONTO	BG
23	SAP SE	SAP	DE
24	Universität Wien (University of Vienna)	UVIE	AT
25	Universiteit Antwerpen (University of Antwerp)	UANTW	BE
26	Institute for Bulgarian Language “Prof. Lyubomir Andreychin”	IBL	BG
27	Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)	FFZG	HR
28	København's Universitet (University of Copenhagen)	UCPH	DK
29	Tartu Ülikool (University of Tartu)	UTART	EE
30	Helsingin Yliopisto (University of Helsinki)	UHEL	FI
31	Centre National de la Recherche Scientifique	CNRS	FR
32	Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)	NYTK	HU
33	Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)	SAM	IS
34	Fondazione Bruno Kessler	FBK	IT
35	Latvijas Universitātes Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia)	IMCS	LV
36	Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)	LKI	LT
37	Luxembourg Institute of Science and Technology	LIST	LU
38	Università ta Malta (University of Malta)	UM	MT
39	Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute)	INT	NL
40	Språkrådet (Language Council of Norway)	LCNOR	NO
41	Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)	IPIPAN	PL
42	Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)	FCULisbon	PT
43	Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)	ICIA	RO
44	University of Cyprus, French and European Studies	UCY	CY
45	Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)	JULS	SK
46	Institut Jožef Stefan (Jozef Stefan Institute)	JSI	SI
47	Centro Nacional de Supercomputación (Barcelona Supercomputing Center)	BSC	ES
48	Kungliga Tekniska högskolan (Royal Institute of Technology)	KTH	SE
49	Universität Zürich (University of Zurich)	UZH	CH
50	University of Sheffield	USFD	UK
51	Universidad de Vigo (University of Vigo)	UVIGO	ES
52	Bangor University	BNGR	UK

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>The West Frisian Language in the Digital Age</b>	<b>3</b>
2.1	General Facts . . . . .	3
2.2	West Frisian in the Digital Sphere . . . . .	6
<b>3</b>	<b>What is Language Technology?</b>	<b>7</b>
<b>4</b>	<b>Language Technology for West Frisian</b>	<b>9</b>
4.1	Language Data . . . . .	9
4.2	Language Technologies and Tools . . . . .	11
4.3	Projects, Initiatives, Stakeholders . . . . .	13
<b>5</b>	<b>Cross-Language Comparison</b>	<b>13</b>
5.1	Dimensions and Types of Resources . . . . .	14
5.2	Levels of Technology Support . . . . .	14
5.3	European Language Grid as Ground Truth . . . . .	15
5.4	Results and Findings . . . . .	15
<b>6</b>	<b>Summary and Conclusions</b>	<b>18</b>

## List of Figures

- |   |   |    |
|---|---|----|
| 1 | Map of the varieties of West Frisian in the province of Fryslân, the Netherlands (Source: Versloot-kartografy 1997) . . . . . | 4  |
| 2 | Overall state of technology support for selected European languages (2022) . .  | 17 |

## List of Tables

- |   |   |    |
|---|---|----|
| 1 | State of technology support, in 2022, for selected European languages with regard to core Language Technology areas and data types as well as overall level of support (light yellow: weak/no support; yellow: fragmentary support; light green: moderate support; green: good support) . . . . . | 16 |
|---|---|----|

## List of Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
BHC	Boarnsterhim Corpus
CL	Computational Linguistics
DLE	Digital Language Equality
ECRML	European Charter for Regional or Minority Languages
ELE	European Language Equality ( <i>this project</i> )
ELE Programme	European Language Equality Programme ( <i>the long-term, large-scale funding programme specified by the ELE project</i> )
ELG	European Language Grid (EU project, 2019-2022)
EU	European Union
FAME	Frisian Audio Mining Enterprise
FCPNM	Framework Convention for the Protection of National Minorities
fyWaC	Frisian Web Corpus
HCI	Human Computer Interaction
KSF	Corpus of Spoken Frisian
LR	Language Resources/Resources
LT	Language Technology/Technologies
ML	Machine Learning
NLP	Natural Language Processing
NLU	Natural Language Understanding
NLG	Natural Language Generation

## Abstract

In this report, we provide an overview of existing Language Technology (LT) for West Frisian, with reference to the language data, tools, and services listed in the European Language Grid.<sup>1</sup> Since LT is central in our everyday lives, such as in spell checkers, search engines, translation software, and virtual assistant technology, it is essential that not only national or majority languages, but also regional or minority languages enjoy digital equality and a high level of digital vitality – a core component of language maintenance and revitalisation. At risk of not achieving digital language equality is West Frisian, an autochthonous minority language spoken in the officially bilingual province of Fryslân (Friesland) in the north of the Netherlands. In an effort to improve the LT situation for West Frisian, we present an analysis of the current data and resources, as well as identify the gaps and challenges that require urgent attention to ensure West Frisian is not negatively affected by digital diglossia (i. e., its exclusion from digital contexts in favour of Dutch, the national language).

Our analysis of current LT resources for West Frisian revealed that there have been significant developments regarding machine translation, spell checkers, monolingual and multimodal corpora, and lexical resources in particular. Crowdsourcing efforts for initiatives such as Mozilla Common Voice have been relatively successful in collecting data that can be utilised in future voice technology projects for West Frisian. There is, however, a lack of bilingual and multilingual text corpora (parallel data), language models, computational grammars, human-computer interaction services, and social media data, among others, as well as of transcribed materials that can be used for the development of speech recognisers and other resources.

Furthermore, although there have been LT programmes for Dutch in recent years, such as the STEVIN programme, regional or minority languages in the Netherlands (e. g., West Frisian) have thus far not received any attention in these projects. There is also no journal or event dedicated specifically to West Frisian LT or artificial intelligence, whereas these are available for Dutch. Moreover, we identified gaps and challenges that highlight the need for increased funding and improved training and retention of skilled LT developers in the province of Fryslân. Overall, despite the impressive progress that has been made to date, there is an evident lack of West Frisian LT resources compared to Dutch and other (official EU) languages.

To improve West Frisian's digital vitality and LT resources, we recommend, among others, that (a) a centre or network should be created by core LT stakeholders in Fryslân to safeguard West Frisian in the digital age, (b) focus should now shift towards collecting data for the development of more advanced tools and services, such as automatic subtitling and voice recognition software, as well as screen readers for visually impaired West Frisian speakers, and (c) West Frisian LT stakeholders should develop initiatives to raise younger generations' awareness of the resources available and to increase the opportunities to use West Frisian LT in everyday life, with the aim of reducing the impact of digital diglossia. We further propose that our recommendations should be integrated into a long-term, national digital language strategy that seeks to facilitate the development of more (multilingual) LT resources, enhance the quality of current West Frisian LT, and enable improved cooperation between the main LT stakeholders in the province of Fryslân, which will collectively contribute to achieving digital language equality for West Frisian.

---

<sup>1</sup> <https://www.european-language-grid.eu>

## Gearfetting

Yn dit rapport jouwe wy in oersjoch fan besteande taaltechnology (LT) foar it Frysk, mei it each op de taaldata, -middels en -tsjinsten dy't yn de European Language Grid steane.<sup>2</sup> Om't LT sintraal stiet yn ús deistich libben, lykas yn staveringshifkers, sykmasines, oersetsoftware en firtuele assistintetechnology, is it essinsjeel dat net allinnich nasjonale of mearderheidstalen, mar ek regionale of minderheidstalen genietsje fan digitale lykweardigens en in heech nivo fan digitale fitaliteit – in kearnkomponint fan taalbehâld en -revitalisaasje. It Frysk hat it risiko om dy digitale taallykweardigens net te berikken. Mei it stribjen om de LT-situaasje foar it Frysk te ferbetterjen, presintearje wy in analyze fan de hjoeddeistige data en middels en identifisearje wy de hiaten en útdagings dy't driuwend it omtinken freegje om derfoar te soargjen dat it Frysk net negatyf beynfloede wurdt troch digitale diglossia (d.w.s. de útsluting fan it Frysk yn digitale fermiddens te'n geunste fan it Nederlânsk).

Ut ús analyze fan hjoeddeistige LT-middels foar it Frysk docht bliken dat der wichtige ûntjouwings west hawwe oangeande masine-oersetting, staveringshifkers, ientalige en multymodale korpora – en yn it bysûnder leksikale middels – en dat crowdsourcing-ynset foar inisjativen lykas Mozilla Common Voice relatyf suksesfol binne yn it sammeljen fan brûkbere data foar takomstige stimtechnologyprojekten foar it Frysk. Der is lykwols in tekoart oan, om in pear te neamen, twatalige en meartalige tekstkorpora (parallele data), taalmodellen, komputasjonele grammatikamodellen, tsjinsten foar ynteraksje tusken minsk en kompjûter, sosjale mediadata, as ek transkribearre materialen dy't brûkt wurde kinne foar de ûntwikkeling fan spraakwerkers en oare middels.

Nettsjinsteande dat de lêste jierren der LT-programma's foar it Nederlânsk west hawwe, lykas it STEVIN-programma, hawwe de streek- of minderheidstalen yn Nederlân oant no ta gjin omtinken krigen yn dy projekten. Der is ek gjin tydskrift of evenemint dat spesifyk wijd is oan Fryske LT of keunstmjittige yntelliginsje, wylst dy der wol binne foar de Nederlânske taal. Boppedat identifisearje wy hiaten en útdagings dy't de needsaak toane foar in ferheging fan finansiering en in ferbettering fan it oplieden en oanhâlden fan betûfte LT-ûntwikkelders yn de provinsje Fryslân. Oer it algemien is der, nettsjinsteande de ûntsachlike foarútgong dy't oant no ta makke is, in dúdlik gebrek oan Fryske LT-middels yn ferliking mei it Nederlânsk en oare (offisjele EU-)talen.

Om de digitale fitaliteit en LT-middels fan it Frysk te ferbetterjen riede wy ûnder oaren oan (a) dat in sintrum of netwurk opset wurde moat fan LT-kearnpartijen yn Fryslân om it Frysk yn it digitale tiidrek te beskermjen, (b) dat der no in ferskowing komme moat nei it sammeljen fan data foar it ûntwikkeljen fan en foar oare middels en tsjinsten lykas automatyske ûndertiteling en spraakwerkningssoftware, lykas skermlêzers foar fisueel beheinde Frysktaligen, en (c) dat belanghawwers fan de Fryske LT inisjativen ûntwikkelje moatte om jongere generaasjes bewust te meitsjen fan de Fryske middels en om de mooglikheden te fergrutsjen om Fryske LT yn it deistich libben te brûken, mei it doel de ynfloed fan digitale diglossia te ferminderjen. Fierders stelle wy út dat ús oanbefellings yntegrearre wurde moatte yn in lanlike digitale taalstrategy op lange termyn, dy't besiket de ûntwikkeling fan mear (meartalige) LT-middels te fasilitearjen, de kwaliteit fan de hjoeddeistige Fryske LT te ferbetterjen en in bettere gearwurking mooglik te meitsjen tusken de wichtichste LT-belanghawwers yn de provinsje Fryslân, dy't kollektyf bydrage sil oan it realisearjen fan digitale taallykweardigens foar it Frysk.

<sup>2</sup> <https://www.european-language-grid.eu>

## 1 Introduction

This study is part of a series that reports on the results of an investigation of the level of support the European languages receive through technology. It is addressed to decision makers at the European and national/regional levels, language communities, journalists, etc. and it seeks to not only delineate the current state of affairs for each of the European languages covered in this series, but to additionally – and most importantly – identify the gaps and factors that hinder further development of research and technology. Identifying such weaknesses will lay the grounds for a comprehensive, evidence-based, proposal of required measures for achieving Digital Language Equality in Europe by 2030.

To this end, more than 40 research partners, experts in more than 30 European languages have conducted an enormous and exhaustive data collection procedure that provided a detailed, empirical and dynamic map of technology support for our languages.<sup>3</sup>

The report has been developed in the frame of the European Language Equality (ELE) project. With a large and all-encompassing consortium consisting of 52 partners covering all European countries, research and industry and all major pan-European initiatives, the ELE project develops a strategic research, innovation, and implementation agenda as well as a roadmap for achieving full digital language equality in Europe by 2030.

## 2 The West Frisian Language in the Digital Age

### 2.1 General Facts

West Frisian (hereinafter referred to as Frisian), or *Frysk* in Frisian, is an autochthonous minority language spoken in *Fryslân* (Friesland), a province in the north of the Netherlands. The official language of the Netherlands is Dutch, and Frisian and Dutch Sign Language are recognised languages (Rijksoverheid, n.d.). In the province of Fryslân, both Dutch and Frisian are official languages. The Netherlands ratified the European Charter for Minority or Regional Languages (ECRML; Council of Europe, 1992) in 1996 in which Frisian is a recognised language under Part III. Moreover, the Dutch government ratified the Framework Convention for the Protection of National Minorities (FCPNM) (Eerste Kamer der Staten-Generaal, n.d.) in 2005, recognising the Frisians as a national minority. The *Wet gebruik Friese taal* (Use of the Frisian Language Act) 2014 affirms both Dutch and Frisian as official languages of the province of Fryslân and reinforces the use and equal position of Frisian (Wettenbank, 2022a). Education of and in the language is documented in the Dutch educational laws as well as in the ECRML, under Parts II and III. Frisian can be used in education, court, and the media, with administrative authorities and public services, and in cultural, social, and economic life. Despite the legal integration of Frisian, however, the use of the language is not guaranteed. For example, Frisian language policy has weakened in recent municipal amalgamations (Schukking and Klinkenberg, 2018), and the use of Frisian in court is under threat, as there have been no Frisian court interpreters available since September 2021 (Anker & Anker, 2021; Omrop Fryslân, 2021a).

In 2021, the Netherlands had over 17.5 million inhabitants (CBS, 2022), and the province of Fryslân had just over 650,000 inhabitants (Centraal Bureau voor de Statistiek, 2022). A periodic sociolinguistic survey commissioned by the provincial government of Fryslân (hereinafter referred to as the Province of Fryslân) revealed that in 2019, 93.1% of the inhabitants of Fryslân reported that they could understand Frisian, 64.1% could speak the language,

<sup>3</sup> The results of this data collection procedure have been integrated into the European Language Grid so that they can be discovered, browsed and further investigated by means of comparative visualisations across languages.



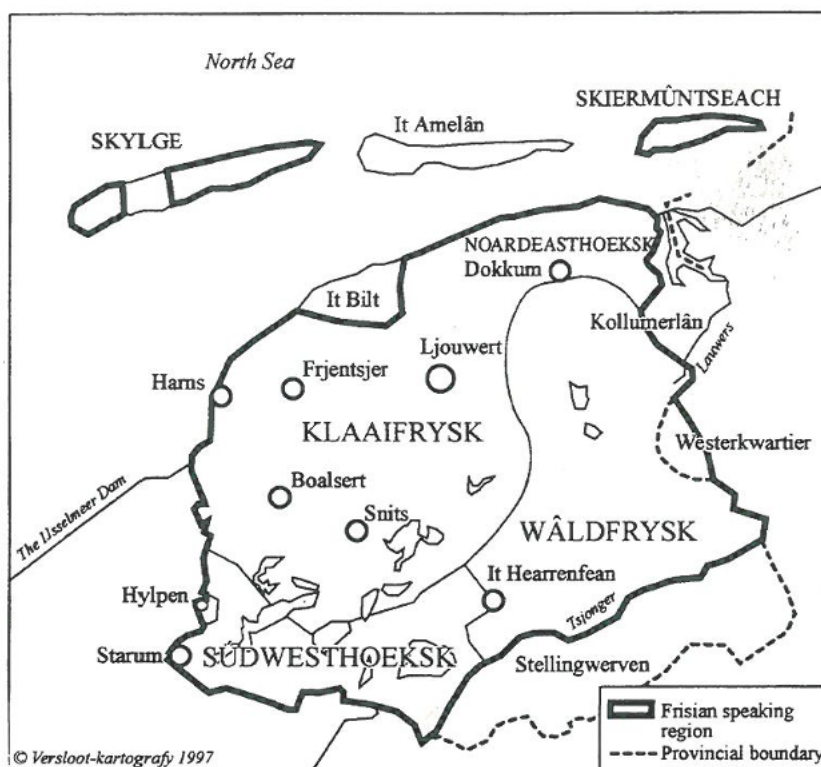


Figure 1: Map of the varieties of West Frisian in the province of Fryslân, the Netherlands (Source: Versloot-kartografy 1997)

52.4% could read Frisian, and 15.9% could write it well (Provinsje Fryslân, 2020a), indicating that Frisian has a reasonable level of vitality in the province. Nevertheless, according to the *UNESCO Atlas of the World's Languages in Danger*, Frisian is considered “vulnerable” (Moseley, 2010).

In the province of Fryslân, there are three main Frisian dialects, namely *Klaaifrysk* (Clay Frisian), *Wâldfrysk* (Forest Frisian), and *Sûdwesthoeksk* (Southwestern variety), as well as a number of smaller ones (e. g., *Skiermûntseagersk*, a variety spoken on the island Schiermonnikoog), and several unique Dutch-Frisian contact varieties, such as *Stedsfrysk* (de Jong and Hoekstra, 2020). Figure 1 displays the varieties of the Frisian language on a map of Fryslân.

Inflection in Frisian is concentrated on the major, open categories of noun, adjective, and verb, with the verb category showing a relatively complex system (Dyk, 2020a). Verb inflection differs from Dutch, which needs to be taken into account when Frisian LT is developed based on a Dutch system. Word formation in Frisian takes place by means of compounding, derivation, and conversion (Dyk, 2020b). A compound word is formed by combining two or more free morphemes. Compounds in Frisian (and other Germanic languages such as Dutch, Danish, and German) are written as a single word (e. g., “language technology”, which is *taaltechnology* in Frisian, and “sewage treatment plant”, which is translated as *riowelwettersuveringsynstallaasje*); this should therefore be considered when developing Frisian LT. A derived word usually consists of one free morpheme and one or more affix(es). In derivation, a free morpheme changes grammatical category without the help of morphological material (Dyk, 2020b). Regarding Frisian syntax, the participle precedes the auxiliary verb clusters, which is the opposite of Dutch (Hoekstra, 2020a). Frisian displays several particular syntax

phenomena, such as the Imperativus-pro-Infinitive construction, or verb ellipsis (Hoekstra, 2020b). This again differs from Dutch and needs to be taken into consideration when LT for Frisian is developed based on Dutch resources.

Frisian has a standardised writing system, which has been under the authority of the Province of Fryslân since 1969. The Province of Fryslân has commissioned the *Fryske Akademy* (Frisian Academy)<sup>4</sup> to develop and describe the system. The last adaptation was in 2014 (Fryslân, 2015). The Latin script is used for the language's writing system, with an alphabet similar to English, though excluding q and x, and including the following vowels with diacritics: â, ê, ô, û, and ú.

Frisian is taught as a mandatory lesson in primary and secondary education within the province of Fryslân. In pre-school education, Frisian can be used as a medium of instruction. There are around 220 pre-school locations with a bilingual (Dutch-Frisian) certificate (Sintrum Frysktalige Berne-opfang, n.d.). In primary education, Frisian is an obligatory school subject and an approved medium of instruction. There were 59 (14.3%) trilingual primary schools (Dutch, Frisian, English) in Fryslân in 2018 (Varkevisser and Walsweer, 2018). In 2018, 67.5% of primary schools in Fryslân did not meet the officially set learning goals for Frisian (Varkevisser and Walsweer, 2018). Frisian is an obligatory school subject in the lower grades of secondary education and it is an optional subject and exam subject in upper secondary education. However, Frisian is not present at all secondary schools, and not all schools teach it in accordance with the set learning goals (Varkevisser and Walsweer, 2018). Moreover, Frisian is offered as an exam subject at only 15 out of 71 secondary school locations (Varkevisser and Walsweer, 2018). In 2019, the Province of Fryslân launched *Taalplan Frysk 2030* (Frisian Language Plan 2030),<sup>5</sup> which aims to ensure that primary and secondary schools teach Frisian at the officially set level by 2030.

Regarding Frisian in vocational education, no undertakings have been signed in Part III of the ECRML. In 2021, however, the Province of Fryslân, the Afûk (a Frisian educational institution and publishing house), and three vocational education institutions agreed to strengthen education in and about Frisian as part of *Taalplan Frysk 2030* (Afûk, 2021). In higher education, there is one Frisian teacher training programme in Fryslân, and the University of Groningen offers a master's programme for postgraduate teacher training for Frisian. At university level, Frisian can be chosen as an optional subject or minor at the universities of Groningen and Amsterdam in the Netherlands, and the universities of Kiel and Flensburg in Germany. Frisian is offered as a bachelor's and a master's programme at Kiel only, as the bachelor's programme that was taught at the University of Groningen was merged into the broader study of Minorities and Multilingualism with an optional Frisian track in 2013. The previously mentioned *Fryske Akademy* conducts scientific research on the Frisian language, as well as the multilingual society and history of Fryslân.

The Frisian broadcaster *Omrop Fryslân*<sup>6</sup> provides news, television, radio, and podcasts in the Frisian language. Online news in Frisian is published by Omrop Fryslân and the website *It Nijs*.<sup>7</sup> The latter is provided by the *Ried fan de Fryske Beweging* (Council of the Frisian Movement – a non-governmental organisation that works to preserve the Frisian language and culture in Fryslân).<sup>8</sup> Smaller local broadcasters and newspapers also publish content in Frisian. The two main newspapers in Fryslân have Dutch as the primary written language, but include content in Frisian to a small extent. There is a steady Frisian literary tradition and there are Frisian literary magazines, such as the Frisian *Ensafh*<sup>9</sup> and the Dutch-Frisian *de*

<sup>4</sup> <https://www.fryske-akademy.nl>

<sup>5</sup> <https://taalplan.frl>

<sup>6</sup> <https://www.omropfryslan.nl>

<sup>7</sup> <https://www.itnijs.frl>

<sup>8</sup> <https://www.fryskebeweging.frl>

<sup>9</sup> <https://www.ensafh.nl>

*Moanne*,<sup>10</sup> which publish in print and online. Frisian theatre is widespread, with two major theatre companies, namely *Tryater*<sup>11</sup> and *Pier21*,<sup>12</sup> as well as popular open air theatre plays, such as in the village of Jorwert. There are also popular musicians and bands who play in Frisian (Jensma, 2015), and (music) festivals or village fêtes. In economic life, a recent study revealed that Frisian was spoken often or very often amongst colleagues and with customers in more than half of the sampled companies (Nicolai et al., 2020). Although Fryslân is an officially bilingual province, and Frisian is actively used in public life, the linguistic landscape (see Landry and Bourhis (1997) for a definition of “linguistic landscape”) of Fryslân does not reflect the situation (Kuipers-Zandberg and Kircher, 2020).

## 2.2 West Frisian in the Digital Sphere

In the Netherlands, 97% of the population has internet access and 86.7% uses it on a daily basis (Centraal Bureau voor de Statistiek, 2021). There are, however, no provincial-level data available regarding internet access. The Fryslân-based top level domain .frl became publicly available on 14 April 2015, with the first site launched on 2 September 2014.<sup>13</sup> As of 31 December 2021, there were approximately 12,092 websites registered with this domain.<sup>14</sup> The .frl domain has become less popular in recent years, as, for instance, registrations were at around 14,558 on 31 December 2018. There is a wide range of Frisian-language content online (on both .nl and .frl sites), covering, for example, political organisations, sport, businesses, news, organisations for the Frisian language, education, writers, poets, and bloggers, tourism, and student associations. All websites (both .frl and .nl) with Frisian-language content that are currently known to the Council of the Frisian Movement are listed on a website<sup>15</sup> supported by the Province of Fryslân. The website also contains background information about Fryslân, the Frisian people, and the Frisian language.

Launched on 2 September 2002, the Frisian Wikipedia (Frisian: *Frysks talige Wikipedy*) has grown considerably.<sup>16</sup> As of December 2021, there were more than 47,500 articles (up from 24 in October 2002) consisting of 15,860,721 words, ranking 99th in terms of number of articles (for comparison, Dutch ranked 6th with 2,078,535 articles according to Wikimedia), and around 41,803 registered users.

The Frisian language is somewhat present on various social media platforms, such as Facebook, Twitter, Instagram, Snapchat, and WhatsApp. A study on the use of Frisian on social media by first language Frisian teenagers (Jongbloed-Faber et al., 2016) revealed that Frisian is used most on WhatsApp (87% used it to some extent), and it is only sometimes used in Facebook status updates (50% reported never using Frisian in Facebook status updates). On Instagram, Facebook, and Twitter, most of the sampled teenagers reported using Dutch, the majority language, more often than Frisian. Overall, it was concluded that the sample of teenagers used Dutch considerably more frequently than Frisian (Jongbloed-Faber, 2021).

An earlier study by Jongbloed-Faber (2015) showed that several factors influence the use of the Frisian language on social media by Frisian speakers aged 15–94. The main factors were the person’s writing skills, number of Frisian contacts on the platform, and language attitude towards Frisian, as well as towards Dutch. Furthermore, people were more likely to react in Frisian to a Frisian post than to post in Frisian. Since the study was conducted, tools to support writing in Frisian, such as Swiftkey, Gboard, online dictionaries, and machine

<sup>10</sup> <https://www.demoanne.nl>

<sup>11</sup> <https://tryater.com>

<sup>12</sup> <https://www.pier21.nl/frl/>

<sup>13</sup> <https://nic.frl>

<sup>14</sup> <https://ntldstats.com/tld/frl>

<sup>15</sup> <https://www.startside.frl>

<sup>16</sup> <https://fy.wikipedia.org/wiki/Haadside>

translation software, have been made available. Further research is, however, needed to investigate the effects of such tools on the use of Frisian.

Among the social media platforms, only Facebook includes Frisian in the language settings. Nevertheless, there is room for improvement, as more English than Frisian is used on the Frisian interface and the option to translate posts is not available. WhatsApp adopts the phone's language setting, but Frisian is not included in the operating systems Android or iOS. It is possible, however, to download Frisian writing support in the form of Swiftkey (2016 for Android and 2017 for iOS) and Gboard (2017). With the help of Frisian community efforts and crowdsourcing events to translate, check, and thus add words, phrases, and sentences to the platform, Frisian has been included in Google Translate since 2016 (Fryske Akademy, 2016; Omrop Fryslân, 2015). There have also been community efforts to include Frisian in Mozilla Common Voice, and Frisian speakers have been asked to record and check Frisian sentences for the platform. In January 2022, there were 49 hours of validated speech out of a total of 125 hours, an increase from the 29 validated hours (out of 75) in 2021. However, more recordings are needed (Fryske Akademy, 2021; Omrop Fryslân, 2021b). It is noted by Jongbloed-Faber (2021) that Frisian is not automatically included by large tech companies, and most efforts to have Frisian included depend on private initiatives and financial support from the provincial government. The data collected for this report support that finding.

### 3 What is Language Technology?

Natural language<sup>17</sup> is the most common and versatile way for humans to convey information. We use language, our natural means of communication, to encode, store, transmit, share and process information. Processing language is a non-trivial, intrinsically complex task, as language is subject to multiple interpretations (ambiguity), and its decoding requires knowledge about the context and the world, while in tandem language can elegantly use different representations to denote the same meaning (variation).

The computational processing of human languages has been established as a specialised field known as *Computational Linguistics* (CL), *Natural Language Processing* (NLP) or, more generally, *Language Technology* (LT). While there are differences in focus and orientation, since CL is more informed by linguistics and NLP by computer science, LT is a more neutral term. In fact, LT is largely multidisciplinary in nature; it combines linguistics, computer science (and notably AI), mathematics and psychology among others. In practice, these communities work closely together, combining methods and approaches inspired by both, together making up *language-centric AI*.

**Language Technology is the multidisciplinary scientific and technological field that is concerned with studying and developing systems capable of processing, analysing, producing and understanding human languages, whether they are written, spoken or embodied.**

With its starting point in the 1950s with Turing's renowned intelligent machine (Turing, 1950) and Chomsky's generative grammar (Chomsky, 1957), LT enjoyed its first boost in the 1990s. This period was signalled by intense efforts to create wide-coverage linguistic resources, such as annotated corpora, thesauri, etc. which were manually labelled for various linguistic phenomena and used to elicit machine readable rules which dictated how language can be automatically analysed and/or produced. Gradually, with the evolution and advances in machine learning, rule-based systems have been displaced by data-based ones, i. e., systems that learn implicitly from examples. In the recent decade of 2010s we observed

<sup>17</sup> This section has been provided by the editors. It is an adapted summary of Agerri et al. (2021) and of Sections 1 and 2 of Aldabe et al. (2021).

a radical technological change in NLP: the use of multilayer neural networks able to solve various sequential labelling problems. The success of this approach lies in the ability of neural networks to learn continuous vector representations of the words (or word embeddings) using vast amounts of unlabelled data and using only some labelled data for fine-tuning.

In recent years, the LT community has been witnessing the emergence of powerful new deep learning techniques and tools that are revolutionising the way in which LT tasks are approached. We are gradually moving from a methodology in which a pipeline of multiple modules was the typical way to implement LT solutions, to architectures based on complex neural networks trained with vast amounts of data, be it text, audio or multimodal. The success in these areas of AI has been possible because of the conjunction of four different research trends: 1) mature deep neural network technology, 2) large amounts of data (and for NLP processing large and diverse multilingual data), 3) increase in high performance computing (HPC) power in the form of GPUs, and 4) application of simple but effective self-learning approaches.

LT is trying to provide solutions for the following main application areas:

- **Text Analysis** which aims at identifying and labelling the linguistic information underlying any text in natural language. This includes the recognition of word, phrase, sentence and section boundaries, recognition of morphological features of words, of syntactic and semantic roles as well as capturing the relations that link text constituents together.
- **Speech processing** aims at allowing humans to communicate with electronic devices through voice. Some of the main areas in Speech Technology are Text to Speech Synthesis, i. e., the generation of speech given a piece of text, Automatic Speech Recognition, i. e., the conversion of speech signal into text, and Speaker Recognition (SR).
- **Machine Translation**, i. e., the automatic translation from one natural language into another.
- **Information Extraction and Information Retrieval** which aim at extracting structured information from unstructured documents, finding appropriate pieces of information in large collections of unstructured material, such as the internet, and providing the documents or text snippets that include the answer to a user's query.
- **Natural Language Generation (NLG)**. NLG is the task of automatically generating texts. Summarisation, i. e., the generation of a summary, the generation of paraphrases, text re-writing, simplification and generation of questions are some example applications of NLG.
- **Human-Computer Interaction** which aims at developing systems that allow the user to converse with computers using natural language (text, speech and non-verbal communication signals, such as gestures and facial expressions). A very popular application within this area are conversational agents (better known as chatbots).

LT is already fused in our everyday lives. As individual users we may be using it without even realising it, when we check our texts for spelling errors, when we use internet search engines or when we call our bank to perform a transaction. It is an important, but often invisible, ingredient of applications that cut across various sectors and domains. To name just very few, in the *health* domain, LT contributes for instance to the automatic recognition and classification of medical terms or to the diagnosis of speech and cognitive disorders. It is more and more integrated in *educational* settings and applications, for instance for educational content mining, for the automatic assessment of free text answers, for providing feedback to learners and teachers, for the evaluation of pronunciation in a foreign language



and much more. In the *law/legal* domain, LT proves an indispensable component for several tasks, from search, classification and codification of huge legal databases to legal question answering and prediction of court decisions.

The wide scope of LT applications evidences not only that LT is one of the most relevant technologies for society, but also one of the most important AI areas with a fast growing economic impact.<sup>18</sup>

## 4 Language Technology for West Frisian

### 4.1 Language Data

#### Monolingual Corpora

There are currently only two monolingual text corpora available for Frisian, namely, Frisian Corpora<sup>19</sup> and the Frisian Web Corpus (fyWaC).<sup>20</sup> Both corpora are not domain specific.

The Frisian Corpora, prepared by the Fryske Akademy, is an annotated corpus of Frisian texts (Old, Mid, and Modern Frisian) that consists of more than 283 million words. The level of detail regarding annotation varies; for example, Mid Frysk is mostly complete and detailed. The texts are lemmatised and part-of-speech (POS) tagged (tense, person, and number). The corpus is designed for research purposes and is accessible free of charge.

The fyWaC is a raw corpus of Frisian texts collected from the internet, consisting of 3 million words, and was prepared by sketchengine.eu in accordance with standards outlined in Kilgarriff et al. (2010). The corpus does not offer POS tagging, but does include tools for word lists, n-grams, and concordances. Access to the fyWaC is available for a fee.

#### Multimodal Corpora

As of 2020/2021, there were four multimodal corpora available for Frisian, namely the FAME! (Frisian Audio Mining Enterprise)<sup>21</sup> Speech Corpus, the Boarnsterhim Corpus (BHC),<sup>22</sup> the *Korpus Sprutsen Frysk* (Corpus of Spoken Frisian; KSF),<sup>23</sup> and Mozilla Common Voice.<sup>24</sup> The FAME corpus and BHC are bilingual.

The FAME! Speech Corpus is a bilingual corpus of radio broadcasts (from 1966–2015) in Frisian and Dutch that consists of 138,000 minutes of audio (with text) from 309 different speakers (e.g. radio presenters and celebrities). The project was led by Radboud University's Centre for Language and Speech Technology, in collaboration with Omrop Fryslân, the Fryske Akademy, GridLine (a company that specialises in language solutions for Dutch), and Tresoar (the repository of the history of Fryslân), under the Dutch Research Council programme *Creatieve Industrie* (Creative Industry; project number 314-99-119).<sup>25</sup> The recordings, which contain Frisian-only and Dutch-only utterances as well as utterances with inter-

<sup>18</sup> In a recent report from 2021, the global LT market was already valued at USD 9.2 billion in 2019 and is anticipated to grow at an annual rate of 18.4% from 2020 to 2028 (<https://www.globenewswire.com/news-release/2021/03/22/2196622/0/en/Global-Natural-Language-Processing-Market-to-Grow-at-a-CAGR-of-18-4-from-2020-to-2028.html>). A different report from 2021 estimates that amid the COVID-19 crisis, the global market for NLP was at USD 13 billion in the year 2020 and is projected to reach USD 25.7 billion by 2027, growing at an annual rate of 10.3% (<https://www.researchandmarkets.com/reports/3502818/natural-language-processing-nlp-global-market>).

<sup>19</sup> <https://frisian.eu/corpus-frontend/frysk/search/>

<sup>20</sup> <https://www.sketchengine.eu/fywac-frisian-corpus/>

<sup>21</sup> <https://fame.ruhosting.nl/wordpress/>

<sup>22</sup> <https://taalmaterialen.ivdnt.org/download/tstc-boarnsterhimcorpus1-0/>

<sup>23</sup> <https://www1.fa.knaw.nl/ksf.html>

<sup>24</sup> <https://commonvoice.mozilla.org/fy-NL>

<sup>25</sup> <https://www.nwo.nl/en/projects/314-99-119>

and intra-sentential and intra-word code-switching, were manually annotated by two native Frisian speakers, which covered, among others, speaker details including gender, dialect, and name (if known). The recordings are from radio programmes about culture, history, literature, sports, nature, agriculture, politics, society, and languages. The duration of the annotated radio broadcasts totals approximately 18.5 hours. This longitudinal and bilingual corpus is suitable for conducting research on language variation in Frisian, such as with regard to code-switching and formal versus informal speech.<sup>26</sup>

The KSF, prepared by the Fryske Akademy, currently consists of 650,000 words (65 hours) of transcribed speech with audio. In this annotated corpus, every speech file is subdivided into utterances that are linked to speakers. The corpus thus contains the following three sorts of research data: 1) linguistic data (transcribed utterances); 2) speech type data associated with utterances; and 3) speaker data associated with utterances.

The BHC, an annotated corpus with both audio and text prepared by the Fryske Akademy under an NWO Veni grant (number 275-75-10, with funding matched by the Fryske Akademy), consists of 250 hours of speech in both Frisian and Dutch, recorded by the same sample of bilingual speakers. The corpus consists of original recordings from 1982–1984 and a replication study that was recorded 35 years later (2018–2019). The data collection spans speech of four generations, and combines panel and trend data (see Sloos et al. (2018), for further details).

The Mozilla Common Voice project, a speech database for the development of speech technology (under CC0 licence), has, as of 2021, collected 4,500 minutes of downloadable speech through crowdsourcing efforts.

A citizen science project that should also be mentioned is *Stimmen fan Fryslân* (Voices of Fryslân),<sup>27</sup> which has played an important role in collecting spoken language data through its smartphone and web applications, particularly of the varieties of Frisian. The intention is to use the corpus of data in future projects that aim to develop, among others, Frisian speech recognition software.

## Lexical Resources

There are various lexical resources available for Frisian, including the following electronic online dictionaries: the *Frysk Hânwurdboek* (concise Frisian dictionary);<sup>28</sup> online Dutch-Frisian and Frisian-Dutch dictionaries;<sup>29</sup> an online Frisian legal dictionary;<sup>30</sup> and a Frisian Historical Dictionary (descriptions in Dutch) with Frisian from 1800 to 1975.<sup>31</sup> A Frisian-English dictionary is available as PDF,<sup>32</sup> as is a word list with words related to the COVID-19 pandemic that was published in 2020. All these lexical resources were composed and produced by the Fryske Akademy, of which most are licensed under CC BY-SA 4.0. The Province of Fryslân has made financial contributions to the development of most of these resources. Moreover, there have been individual efforts to compose (smaller) dictionaries and word lists.

On the web page *Taalmaterialen* (Language materials)<sup>33</sup> developed by the Dutch Language Institute (Dutch: *Instituut voor de Nederlandse Taal*), which is a rich source of Dutch language

<sup>26</sup> For further information about the FAME corpus, see Yilmaz et al. (2016).

<sup>27</sup> <http://stimmen.nl>

<sup>28</sup> <https://frisian.eu/dictionaries/fhwbapp/index.html>

<sup>29</sup> <https://frisian.eu/TEST/dictionaries/onfwapp/index.html>

<sup>30</sup> <https://frisian.eu/dictionaries/jurwbapp/index.html>

<sup>31</sup> <https://gtb.ivdnt.org/search/>

<sup>32</sup> [https://www.fryske-akademy.nl/fileadmin/inhoud/beelden/homepage/Kennis/Utjeften/Downloads/Frysk-Ingelsk\\_wurdboek\\_2000.pdf](https://www.fryske-akademy.nl/fileadmin/inhoud/beelden/homepage/Kennis/Utjeften/Downloads/Frysk-Ingelsk_wurdboek_2000.pdf)

<sup>33</sup> <https://ivdnt.org/taalmaterialen/>

materials, one Frisian dictionary (Frisian: *Wurdboek fan de Fryske taal* – Frisian Historical Dictionary) is listed.<sup>34 35</sup>

## 4.2 Language Technologies and Tools

There are several tools and services available for Frisian, though not in abundance. Most are developed by the Fryske Akademy and financially supported by the Province of Fryslân. The following umbrella websites offer these resources or have compiled lists of these resources for easy access: [Frysker.nl](https://frysker.nl)<sup>36</sup> by the Fryske Akademy, for machine translation, dictionaries, a spell checker, and other resources; [Frisian.eu](https://frisian.eu),<sup>37</sup> for tools and services developed by the Fryske Akademy; [Taalhelp.frl](https://taalhelp.frl)<sup>38</sup> by the Afûk, for an overview of online dictionaries, apps, plug-ins, and games; and [Taalportaal.org](https://taalportaal.org),<sup>39</sup> which provides a comprehensive and authoritative scientific grammar for Frisian, as well as for Dutch and Afrikaans.

### Text and Speech Analysis

There are some tools and services developed by the Fryske Akademy, in collaboration with other organisations, for text and speech analysis regarding Frisian.

UDPipe Frysk is an online service for tokenising, lemmatising, POS-tagging, and dependency parsing Frisian texts, developed by the Fryske Akademy and the University of Groningen within a CLARIAH-Plus project financed by the Dutch Research Council (grant number 184.034.023).<sup>40</sup> The service, which is still under development, is available free of charge and is licensed under CC BY-NC-SA.

Visible Vowels is a web application for the analysis, normalisation, and visualisation of acoustic vowel measurements (Heeringa and Van de Velde, 2017),<sup>41</sup> and Grapheme to Phoneme can be used for retrieving phonemes from Frisian texts.<sup>42</sup> GraphQL Language API for Frisian-Dutch allows searches for lemma, text, information, source, and translation.<sup>43</sup> TEI ODDs (Text Encoding Initiative - One Document Does it all) for dictionaries<sup>44</sup> and TEI ODDs for corpora<sup>45</sup> support the editorial work of the Fryske Akademy with TEI resources.

### Speech Processing

Based on the aforementioned project FAME! (see Section 4.1), further developments are expected, including a speech recogniser that automatically transcribes Frisian and Dutch speech simultaneously, which is currently work in progress. Such developments are essential, as it is mandatory to provide Dutch subtitles for provincial, municipality, and water council meetings, according to the Dutch law on the accessibility of the government (Wettenbank, 2022b) and the European directive on the accessibility of the websites and mobile applications of public sector bodies (European Parliament & Council of the European Union, 2016). A grant was awarded in late 2020 to further develop the FAME! speech recogniser for this

<sup>34</sup> <https://taalmaterialen.ivdnt.org/download/pp-woordenboek-der-friese-taal-j/>

<sup>35</sup> <https://gtb.ivdnt.org/search/>

<sup>36</sup> <https://frysker.nl>

<sup>37</sup> <https://frisian.eu>

<sup>38</sup> <https://taalhelp.frl>

<sup>39</sup> <https://www.taalportaal.org/taalportaal/topic/>

<sup>40</sup> <https://frisian.eu/udpipeapp/>

<sup>41</sup> <https://www.visiblevowels.org>

<sup>42</sup> <https://frisian.eu/graph2phonapp/>

<sup>43</sup> <https://frisian.eu/languageapidocs/>

<sup>44</sup> [https://bitbucket.org/fryske-akademy/online-dictionaries/src/master/src/main/resources/odd/tei\\_dictionaries.odd](https://bitbucket.org/fryske-akademy/online-dictionaries/src/master/src/main/resources/odd/tei_dictionaries.odd)

<sup>45</sup> [https://bitbucket.org/fryske-akademy/tei-encoding/src/master/src/main/resources/customization/v2\\_0/](https://bitbucket.org/fryske-akademy/tei-encoding/src/master/src/main/resources/customization/v2_0/)



purpose (Van den Heuvel et al., 2021). However, funding for developing automatic subtitling for Frisian is not available. Financial support in general is also lacking, as there is no policy document that supports speech technology for Frisian (Van de Velde, 2021). Moreover, there is a limited number of transcribed materials to train speech recognisers (Van de Velde, 2021).

A PhD student at the University of Groningen/Campus Fryslân is currently conducting research to develop a Frisian speech synthesiser, and to investigate whether other languages can be used (and to what extent) for a small language that does not have an extensive amount of data available.

For Frisian text-to-speech, there are two main (paid) services, namely *Fluency*<sup>46</sup> and *Readspeaker*.<sup>47</sup> Fluency is a project by Readspeak, the Afûk, and the Fryske Akademy, financed by the Province of Fryslân, which have developed a Frisian speech synthesiser based on a Dutch synthesiser. The final product is available via Fluency and Readspeak. Readspeak has continued the development of a Frisian text-to-speech tool, which is used by some Frisian websites, such as the Frisian news platforms Omrop Fryslân and It nijs.

### Translation Technologies and Authoring Tools

*Frysker*,<sup>48</sup> a platform for Frisian-Dutch and Dutch-Frisian machine translation and dictionaries, and a Frisian spell checker, was launched in 2021. Frysker is a project carried out by the Fryske Akademy, and an editorial staff ensures the quality of the platform. Frysker is financed by the Province of Fryslân, and is developed in collaboration with the following partners: BW H ontwerpers, GridLine, KNAW Humanities Cluster / Maarten van Gompel, and the Afûk.

The neural machine translation tool (*Oersetter*)<sup>49</sup> is of good quality, and an improvement on the earlier version on *Taalweb Frysk* (Language web Frisian), which Frysker has replaced. Currently, translating from Frisian to Dutch can cause issues when the Frisian text is written phonetically (which is often the case, as Frisian writing proficiency is low). The programme has a self-learning element and trains itself with the data provided by users, so improvement is expected.

The Frysker dictionary service allows people to find translations, definitions, synonyms, (dialectal) variants, and sayings. The dictionary service contains the following dictionaries developed by the Fryske Akademy: *Foarkarswurldlist foar it Frysk* (list of preferred spelling, 2015),<sup>50</sup> *it Frysk Hânwurdboek* (concise Frisian dictionary, 2008);<sup>51</sup> a Frisian-Dutch dictionary (1984) and Dutch-Frisian dictionary (1985);<sup>52</sup> a legal dictionary;<sup>53</sup> data from the Online Dutch-Frisian dictionary (2021);<sup>54</sup> and *Paadwizer Fryske Stavering* (Frisian Spelling Guidelines).<sup>55</sup>

Frysker also offers a spell checker.<sup>56</sup> Other Frisian spell checkers are available as add-ons for Mozilla Firefox and Mozilla Thunderbird, as well as for LibreOffice and OpenOffice.

It is also worth noting the Frisian-speaking community's efforts in the development of translation technology. In 2015, a successful campaign was held to include Frisian in Google Translate, during which there were many activities to which Frisian speakers could con-

<sup>46</sup> <https://fluency.nl>

<sup>47</sup> <https://www.readspeaker.com/languages-voices/>

<sup>48</sup> <https://frysker.nl>

<sup>49</sup> <http://frysker.nl/oersetter>

<sup>50</sup> <https://taalweb.frl/foarkarswurldlist>

<sup>51</sup> <https://frisian.eu/dictionary-services/fhwbapp/index.html>

<sup>52</sup> <https://frisian.eu/dictionary-services/nfwbapp/index.html>

<sup>53</sup> <https://frisian.eu/dictionaries/jurwbapp/index.html>

<sup>54</sup> <https://frisian.eu/TEST/dictionary-services/onfwapp/index.html>

<sup>55</sup> <http://frysker.nl/paadwizer/0-foarof>

<sup>56</sup> <https://frysker.nl/hifker/>

tribute. Since 2016, Frisian has been available on the platform (functions: type, handwriting, and image translation).

### Information Extraction and Information Retrieval

For the extraction/retrieval of Frisian data from Twitter, language detectors available in the R programming language have been used, namely the function `textcat` from the `textcat` package, and the function `detect_language` from Google's Compact Language Detector 3 (CLD3) package (Dijkstra et al., 2021).

## 4.3 Projects, Initiatives, Stakeholders

In the Netherlands, there is no dedicated LT programme at the national level (for Dutch or Frisian), but some LT-related funding is available primarily for Dutch (Steurs et al., 2022; Rehm et al., 2020; Van de Velde, 2021). Artificial intelligence strategies have also been developed for Dutch, such as the Dutch Digitalisation Strategy 2.0 (2019)<sup>57</sup> and Strategic Action Plan for AI (2019),<sup>58</sup> though no funding is available for LT development through AI. Frisian is not, however, included in the aforementioned AI strategies, nor are there LT or AI programmes at the provincial level for Frisian. Language technology is also not mentioned in the Frisian provincial policy 2021–2024 for culture, language, heritage, and education (Provinsje Fryslân, 2020b).

There have been programmes in recent years that aimed to develop Dutch LT, often in cooperation with the LT community in Flanders, Belgium, such as the STEVIN programme (Spyns and D'Halleweyn, 2013),<sup>59</sup> which ran between 2004 and 2011 and was coordinated by the Dutch Language Union.<sup>60</sup> However, regional or minority languages in the Netherlands (e. g., Frisian) have not been included in such programmes thus far.

Furthermore, the Netherlands plays a role in the Common Language Research Infrastructure (CLARIN), a European research infrastructure. Resources developed in CLARIN NL and CLARIAH (an infrastructure for arts and humanities that brings together CLARIN and DARIAH) NL projects are stored in the Dutch CLARIN Portal Pages CLAPOP; some Frisian resources are listed on the CLARIN/CLAPOP website.<sup>61</sup>

There is an annual conference and journal on Computational Linguistics in the Netherlands dedicated to the Dutch language,<sup>62</sup> but there are no such events or journals specifically for Frisian.

This report highlights that the Fryske Akademy is a key actor in the development of LT for Frisian, as it provides, among others, lexical resources and corpora, and is involved in numerous LT projects, which are often financed by the Province of Fryslân.

## 5 Cross-Language Comparison

The LT field<sup>63</sup> as a whole has evidenced remarkable progress during the last years. The advent of deep learning and neural networks over the past decade together with the considerable increase in the number and quality of resources for many languages have yielded re-

<sup>57</sup> <https://www.nederlanddigitaal.nl/english/dutch-digitalisation-strategy-2.0>

<sup>58</sup> [https://knowledge4policy.ec.europa.eu/ai-watch/netherlands-ai-strategy-report\\_en](https://knowledge4policy.ec.europa.eu/ai-watch/netherlands-ai-strategy-report_en)

<sup>59</sup> *Spraak- en Taaltechnologische Voorzieningen in het Nederlands* (Speech and Language Technological Resources for Dutch): <https://www.nemokennislink.nl/publicaties/stevin-2004-2010/>

<sup>60</sup> *Nederlandse Taalunie*: <https://taalunie.org/home>

<sup>61</sup> <https://portal.clarin.nl/search/node/frisian>

<sup>62</sup> <https://www.clinjournal.org/clinj>

<sup>63</sup> This section has been provided by the editors.

sults unforeseeable before. However, is this remarkable progress equally evidenced across all languages? To compare the level of technology support across languages, we considered more than 11,500 language technology tools and resources in the catalogue of the European Language Grid platform (as of January 2022).

## 5.1 Dimensions and Types of Resources

The comparative evaluation was performed on various dimensions:

- The current state of technology support, as indicated by the availability of tools and services<sup>64</sup> broadly categorised into a number of core LT application areas:
  - Text processing (e. g. part-of-speech tagging, syntactic parsing)
  - Information extraction and retrieval (e. g. search and information mining)
  - Translation technologies (e. g. machine translation, computer-aided translation)
  - Natural language generation (e. g. text summarisation, simplification)
  - Speech processing (e. g. speech synthesis, speech recognition)
  - Image/video processing (e. g. facial expression recognition)
  - Human-computer interaction (e. g. tools for conversational systems)
- The potential for short- and mid-term development of LT, insofar as this potential can be approximated by the current availability of resources that can be used as training or evaluation data. The availability of data was investigated with regard to a small number of basic types of resources:
  - Text corpora
  - Parallel corpora
  - Multimodal corpora (incl. speech, image, video)
  - Models
  - Lexical resources (incl. dictionaries, wordnets, ontologies etc.)

## 5.2 Levels of Technology Support

We measured the relative technology support for 87 national, regional and minority European languages with regard to each of the dimensions mentioned above based on their respective coverage in the ELG catalogue. For the types of resources and application areas, the respective percentage of resources that support a specific language over the total number of resources of the same type was calculated, as well as their average. Subsequently each language was assigned to one band per resource type and per application area and to an overall band, on a four-point scale, inspired by the scale used in the META-NET White Paper Series, as follows:

1. **Weak or no support:** the language is present (as content, input or output language) in <3% of the ELG resources of the same type
2. **Fragmentary support:** the language is present in  $\geq 3\%$  and <10% of the ELG resources of the same type

<sup>64</sup> Tools tagged as “language independent” without mentioning any specific language are not taken into account. Such tools can certainly be applied to a number of languages, either as readily applicable or following fine-tuning, adaptation, training on language-specific data etc., yet their exact language coverage or readiness is difficult to ascertain.

3. **Moderate support:** the language is present in  $\geq 10\%$  and  $< 30\%$  of the ELG resources of the same type
4. **Good support:** the language is present in  $\geq 30\%$  of the ELG resources of the same type<sup>65</sup>

The overall level of support for a language was calculated based on the average coverage in all dimensions investigated.

### 5.3 European Language Grid as Ground Truth

At the time of writing (January 2022), the ELG catalogue comprises more than 11,500 metadata records, encompassing both data and tools/services, covering almost all European languages – both official and regional/minority ones. The ELG platform harvests several major LR/LT repositories<sup>66</sup> and, on top of that, more than 6,000 additional language resources and tools were identified and documented by language informants in the ELE consortium. These records contain multiple levels of metadata granularity as part of their descriptions.

It should be noted that due to the evolving nature of this extensive catalogue and differing approaches taken in documenting records, certain levels of metadata captured are not yet at the level of consistency required to carry out a reliable cross-lingual comparison at a granular level. For example, information captured on corpora size, annotation type, licensing type, size unit type, and so on, still varies across records for many languages, while numerous gaps exist for others. As the ELG catalogue is continuously growing, the comprehensiveness, accuracy and level of detail of the records will naturally improve over time. Moreover, the Digital Language Equality (DLE) metric will allow for dynamic analyses and calculations of digital readiness, based on the much finer granularity of ELG records as they mature.<sup>67</sup>

For the purposes of high-level comparison in this report, the results presented here are based on relative counts of entries in the ELG for the varying types of data resources and tools/services for each language. As such, the positioning of each language into a specific level of technology support is subject to change and it reflects a snapshot of the available resources on January 2022.

That said, we consider the current status of the ELG repository and the higher level findings below adequately representative with regard to the current existence of LT resources for Europe's languages.

### 5.4 Results and Findings

As discussed above, our analysis takes into account a number of dimensions for data and tools/services. Table 1 reports the detailed results per language per dimension investigated and the classification of each language into an overall level of support.

The best supported language is, as expected, English, the only language that is classified in the *good support* group. French, German and Spanish form a group of languages with *moderate support*. Although they are similar to English in some dimensions (e. g. German in terms of available speech technologies and Spanish in terms of available models), overall they have

<sup>65</sup> The thresholds for defining the four bands were informed by an exploratory *k*-means 4-cluster analysis based on all data per application and resource type, in order to investigate the boundaries of naturally occurring clusters in the data. The boundaries of the clusters (i. e., 3%, 10% and 30%) were then used to define the bands per application area and resource type.

<sup>66</sup> At the time of writing, ELG harvests ELRC-SHARE, LINDAT/CLARIAH-CZ, CLARIN.SI, CLARIN-PL and HuggingFace.

<sup>67</sup> Interactive comparison visualisations of the technology support of Europe's languages will be possible on the ELG website using a dedicated dashboard, which dynamically analyses the resources available in the ELG repository, from the middle of 2022 onwards.



not yet reached the coverage that English has according to the ELG platform. All other official EU languages are clustered in the *fragmentary support* group, with the exception of Irish and Maltese, which have only *weak or no support*. From the remaining languages, (co-)official at national or regional level in at least one European country and other minority and lesser spoken languages,<sup>68</sup> Norwegian and Catalan belong to the group of languages with *fragmentary support*. Basque, Galician, Icelandic and Welsh are borderline cases; while they are grouped in the *fragmentary support* level, they barely pass the threshold from the lowest level. All other languages are supported by technology either weakly or not at all. Figure 2 visualises our findings.

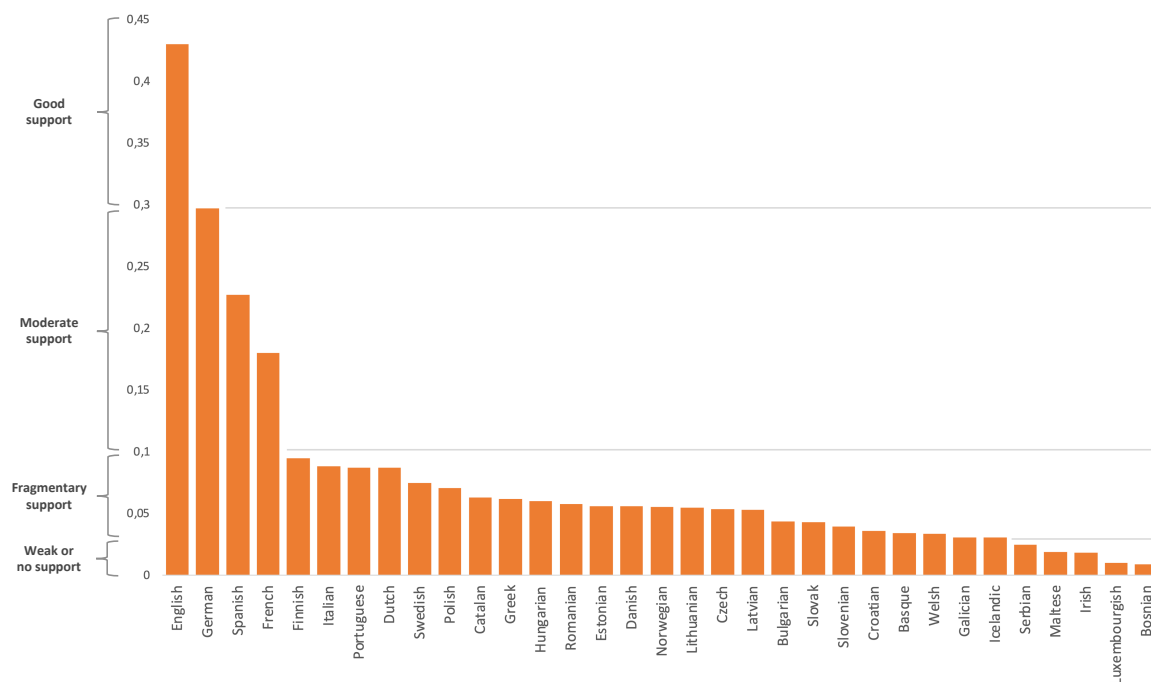


Figure 2: Overall state of technology support for selected European languages (2022)

While a fifth level, *excellent support*, could have been foreseen in addition to the four levels described in Section 5.2, we decided not to consider this level for the grouping of languages. Currently no natural language is optimally supported by technology, i. e., the goal of *Deep Natural Language Understanding* has not been reached yet for any language, not even for English, the best supported language according to our analysis. While recently there have been many breakthroughs in AI, Computer Vision, ML and LT, we are still far from the grand challenge of highly accurate deep language understanding, which is able to seamlessly integrate modalities, situational and linguistic context, general knowledge, meaning, reasoning, emotion, irony, sarcasm, humour, culture, explain itself at request, and be done as required on the fly and at scale. A language can only be considered as excellently supported by technology if and when this goal of Deep Natural language Understanding has been reached.

The results of the present comparative evaluation reflect, in terms of distribution and im-

<sup>68</sup> In addition to the languages listed in Table 1, ELE also investigated Alsatian, Aragonese, Arberesh, Aromanian, Asturian, Breton, Cimbrian, Continental Southern Italian (Neapolitan), Cornish, Eastern Frisian, Emilian, Franco-Provençal (Arpitan), Friulian, Gallo, Griko, Inari Sami, Karelian, Kashubian, Ladin, Latgalian, Ligurian, Lombard, Lower Sorbian, Lule Sami, Mocheno, Northern Frisian, Northern Sami, Picard, Piedmontese, Pite Sami, Romagnol, Romany, Rusyn, Sardinian, Scottish Gaelic, Sicilian, Skolt Sami, Southern Sami, Tatar, Tornedalian Finnish, Venetian, Võro, Walser, Yiddish.



balance, the results of the META-NET White Paper Series (Rehm and Uszkoreit, 2012). The complexities of the analyses clearly differ across 2012 and 2022 studies, and as such, a direct comparison between the two studies can therefore not be made. However, we can instead compare the relative level of progress made for each language in the meantime. It is undebatable that the technology requirements for a language to be considered digitally supported today have changed significantly (e. g. the prevalent use of virtual assistants, chat bots, improved text analytics capabilities, etc.). Yet also the imbalance in distribution across languages still exists.

The results of this analysis are only informative of the relative positioning of languages, but not of the progress achieved within a specific language. The LT field as a whole has significantly progressed in the last ten years and remarkable progress has been achieved for specific languages in terms of quantity, quality and coverage of tools and language resources. Yet, the abysmal distance between the best supported languages and the minimally supported ones is still evidenced in 2022. It is exactly this distance that needs to be ideally eliminated, if not at least reduced, in order to move towards Digital Language Equality and avert the risks of digital extinction.

## 6 Summary and Conclusions

The findings presented in this report illustrate the significant differences in LT availability between European languages. Although there are no national LT programmes for Dutch (the national language in the Netherlands), there are national-level AI strategies. These do not, however, take into account Frisian, or other regional languages (e.g. Low Saxon and Limburgish). Regarding the Frisian language, it is evident from our analysis that there are gaps in, for instance, automatic subtitling and speech recognition software. Nevertheless, there is a relatively strong research community in Fryslân that develops Frisian LT, which is mainly based at the Fryske Akademy and has connections with several Dutch universities, including the universities of Groningen and Amsterdam. Such collaboration has facilitated good knowledge transfer thus far, but there is room for improvement.

Over the past 5 years, there have been substantial advances in machine translation, spell checker software, and online dictionaries for Frisian, and projects such as Mozilla Common Voice have played a key role in collecting data that can be used in the future development of Frisian voice technology. There are, however, currently no bilingual and multilingual text corpora (parallel data), language models, computational grammars, or human-computer interaction services for Frisian. Corpora comprising social media texts are also missing. Such data are particularly difficult to collect partially because of the GDPR regulation considerations. There are also gaps regarding automatic speech recognition, natural language generation, and automatic subtitling, among others. Moreover, there is a lack of transcribed materials that can be used to train speech recognisers and other LT resources. The availability of LT resources is thus still limited compared to Dutch and other languages such as English. This restricts LT and AI developments, particularly those that require large amounts of data (e. g. deep learning methodologies). Furthermore, the general lack of data and funding for Frisian LT development highlights the need for increased funding as well as improved training and retention of skilled LT developers in the province of Fryslân.

In this era of growing globalisation, the limited LT developed for Frisian and the differences in quality of these resources leave the language at risk of being neglected in the efforts to achieve digital language equality. Improved digital language vitality, a core component of language maintenance and revitalisation, is also essential for Frisian, given that it is vulnerable (Moseley, 2010) and the digital presence of a language has a significant influence on its visibility, status, and prestige. Moreover, as LT increasingly becomes part of everyday

life, it also enters the informal (spoken) settings in which Frisian is mainly used. However, if LT for Frisian is lacking, Frisian speakers will shift to Dutch or another majority language (e. g. English) to communicate with systems in day-to-day life and in society (Omrop Fryslân, 2018).

Considering the strengths and weaknesses of Frisian LT, we wish to make several recommendations that require urgent attention. Our recommendations aim to facilitate the development of more (multilingual) LT resources, to improve the quality of existing Frisian LT resources, and to enable better cooperation between key LT stakeholders in the province of Fryslân, contributing to achieving digital equality for Frisian.

- A centre or network should be established by core LT stakeholders in Fryslân that aims to support and enrich education in the fields of LT and AI, increase the visibility and ensure the sustainability of existing and future Frisian LT resources, and facilitate improved knowledge transfer and collaboration between academic and industry stakeholders (e. g. by organising an annual conference). This would contribute to boosting the province's economy, improving the vitality of the Frisian language, and ensuring digital equality for Frisian.
- There have been significant developments in terms of dictionaries, translation software, and spell checkers, but there should be a shift in focus to other technology such as automatic subtitling, voice recognition, screen readers for visually impaired Frisian speakers, domain-specific datasets (e. g. legal and medical), and improved multimedia resources, of which the data and source code should be made available under open source licensing to allow further developments by larger institutions with wider skillsets and more resources. Funding for such projects should be made available by the provincial and national governments to key organisations in Fryslân.
- Frisian LT stakeholders should develop initiatives that aim to raise awareness of LT resources among younger generations. Digital diglossia is a major threat to Frisian (i. e. Frisian is excluded from digital contexts in favour of Dutch, the national language), with younger generations often opting to use Dutch, the language that allows them greater and better access to various content, among others, and to more advanced technologies. Efforts thus need to be made to reduce the generational language gap and increase the opportunities to use Frisian LT in everyday life (e. g., in apps or with virtual assistants), preventing the digital extinction of Frisian, considering, for instance, its limited use on social media (Jongbloed-Faber, 2015, 2021).
- In September 2021, a new MSc in voice technology was launched at the University of Groningen/Campus Fryslân, which is the only master's programme in continental Europe that is exclusively dedicated to voice technology. In the coming years, this master's programme has an important role to play in the development of Frisian LT. It is therefore recommended that this programme provide sufficient opportunities (e. g. internships and jobs, with focus on the Frisian language) in collaboration with other organisations and the Province of Fryslân to ensure that talent is nurtured and retained in the province.
- Although it is expected that Frisian LT will be needed in, for instance, the healthcare sector (e. g., healthcare apps or robots), government accessibility, and consumer products, there is still a lack of knowledge regarding Frisian speakers' demand for Frisian LT. Research should thus be conducted on what resources should be prioritised and developed in future projects.

To improve the overall LT situation for Frisian, the aforementioned recommendations should be combined into a long-term, national digital language strategy, which is currently



absent. It is essential that such a strategy is developed and implemented at not only the provincial level but also at the national level. Frisian is a recognised language in the Netherlands through the ratification of the ECRML and FCPNM, and the national government is obliged to support the use of Frisian in private and public domains, to promote equality, and to create the conditions for preserving and developing the minority language, amongst other markers of cultural identity. A long-term, national strategic plan is therefore vital for safeguarding Frisian in the digital age, increasing support for and enriching LT-related education, and facilitating the extensive data collection and annotation that are still required to advance Frisian LT and ensure its digital vitality and equality.

## References

- Afûk. *Fryske mbû's sette skouders ûnder it Frysk* [frisian intermediate vocational education institutions put their shoulder to the wheel for the frisian language], 2021. URL <https://afuk.frl/fryske-mbus-sette-skouders-under-it-frysk/>.
- Rodrigo Agerri, Eneko Agirre, Itziar Aldabe, Nora Aranberri, Jose Maria Arriola, Aitziber Atutxa, Gorka Azkune, Arantza Casillas, Ainara Estarrona, Aritz Farwell, Iakes Goenaga, Josu Goikoetxea, Koldo Gojenola, Inma Hernaez, Mikel Irukieta, Gorka Labaka, Oier Lopez de Lacalle, Eva Navas, Maite Oronoz, Arantxa Otegi, Alicia Pérez, Olatz Perez de Viñaspre, German Rigau, Jon Sanchez, Ibon Saratzaga, and Aitor Soroa. European Language Equality D1.2: Report on the state of the art in Language Technology and Language-centric AI, September 2021. URL [https://european-language-equality.eu/wp-content/uploads/2021/10/ELE\\_Deliverable\\_D1\\_2.pdf](https://european-language-equality.eu/wp-content/uploads/2021/10/ELE_Deliverable_D1_2.pdf).
- Itziar Aldabe, Georg Rehm, German Rigau, and Andy Way. European Language Equality D3.1: Report on existing strategic documents and projects in LT/AI, November 2021. URL [https://european-language-equality.eu/wp-content/uploads/2021/12/ELE\\_\\_Deliverable\\_D3\\_1\\_revised\\_.pdf](https://european-language-equality.eu/wp-content/uploads/2021/12/ELE__Deliverable_D3_1_revised_.pdf).
- Anker & Anker. *Werkweigering tolken Fries zet zorgvuldige rechtsgang onder druk* [Work strike of Frisian interpreters puts careful court procedure under pressure], 2021. URL [https://www.ankerenanker.nl/nieuws/archief/8185\\_werkweigering-tolken-fries-zet-zorgvuldige-rechtsgang-onder-druk](https://www.ankerenanker.nl/nieuws/archief/8185_werkweigering-tolken-fries-zet-zorgvuldige-rechtsgang-onder-druk).
- CBS. *Bevolkingsontwikkeling; maand en jaar* [population development: month and year], 2022. URL <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83474NED/table?ts=1641915928060>. CBS Statline. Retrieved January 13, 2022.
- Centraal Bureau voor de Statistiek. *Internettoegang en internetactiviteiten; persoonskenmerken* [internet access and activities; personal characteristics], 2021. URL <https://www.cbs.nl/nl-nl/cijfers/detail/84888NED>. Retrieved January 21, 2022.
- Centraal Bureau voor de Statistiek. *Regionale kerncijfers Nederland* [regional core numbers of the netherlands], 2022. URL <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70072NED/table?fromstatweb>. CBS Statline. Retrieved January 31, 2022.
- Noam Chomsky. *Syntactic structures*. The Hague: Mouton, 1957.
- G. de Jong and E Hoekstra. *A general introduction to Frisian*, 2020. URL <https://www.taalportaal.org/taalportaal/topic/pid/topic-14225224491227143>. Taalportaal.
- J. Dijkstra, W. J. Heeringa, L. Jongbloed-Faber, and H. Van de Velde. Using twitter data for the study of language change in low-resource languages. a panel study of relative pronouns in frisian, 2021. URL <https://doi.org/10.3389/frai.2021.644554>. *Frontiers in Artificial Intelligence*.
- S. Dyk. *Inflection*, 2020a. URL <https://www.taalportaal.org/taalportaal/topic/pid/topic-14248805702102347>. Taalportaal.

- S. Dyk. *Word Formation*, 2020b. URL <https://www.taalportaal.org/taalportaal/topic/pid/topic-14303771622988096>. Taalportaal.
- ECRML; Council of Europe. *Text of the European Charter for Regional or Minority Languages*, 1992. URL <https://www.coe.int/en/web/european-charter-regional-or-minority-languages/text-of-the-charter>.
- Eerste Kamer der Staten-Generaal. Verdrag inzake de bescherming van nationale minderheden [convention for the protection of national minorities], n.d. URL [https://www.eerstekamer.nl/wetsvoorstel/26389\\_verdrag\\_inzake\\_de](https://www.eerstekamer.nl/wetsvoorstel/26389_verdrag_inzake_de). Eerste Kamer der Staten-Generaal.
- European Parliament & Council of the European Union. *Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies*. EUR-Lex, 2016. URL <https://eur-lex.europa.eu/eli/dir/2016/2102/oj>.
- Fryske Akademy. *Frysk yn Google Translate*. [Frisian in Google Translate], 2016. URL <https://www.fryske-akademy.nl/nijs-en-aginda/argivearre-nijsberjocht/news/detail/frysk-yn-google-translate/>. Fryske Akademy.
- Fryske Akademy. *20-26 septimber – Fryske ynsprekwike Common Voice*. [20-26 September - Frisian recording week Common Voice], 2021. URL <https://www.fryske-akademy.nl/nijs-en-aginda/nijsberjocht/news/detail/20-26-septimber-fryske-ynsprekwike-common-voice/>. Fryske Akademy.
- Fryslân. *Feroardering fan Deputearre Steaten fan Fryslân fan 27 jannewaris 2015, hâldende regels oangeande de offisjele stavering fan de Fryske taal 2014*. [Ordinance of the Frisian Provincial Executive of 27 January 2015, rules in effect regarding the official spelling of the Frisian language 2014], 2015. URL <https://zoek.officielebekendmakingen.nl/prb-2015-477.html>.
- W. Heeringa and H. Van de Velde. Visible vowels: A tool for the visualization of vowel variation, 2017. URL [https://pure.knaw.nl/ws/portalfiles/portal/8997153/CLARIN2018\\_final.pdf](https://pure.knaw.nl/ws/portalfiles/portal/8997153/CLARIN2018_final.pdf). Proc. Inter-speech 2017, 4034–4035.
- E. Hoekstra. Constructions with verbs and verb clusters, 2020a. URL <https://www.taalportaal.org/taalportaal/topic/pid/topic-14115624524539056>. Taalportaal.
- E. Hoekstra. Ellipsis of ordinary infinitives and to-infinitives, 2020b. URL <https://www.taalportaal.org/taalportaal/topic/pid/topic-14133785397588266>. Taalportaal.
- Goffe Jensma. Friesland en rowwen hêze. In *Het dorp & de wereld: Over dertig jaar Rowwen Hêze*, pages 116–229. Uitgeverij Vantilt, 2015.
- L. Jongbloed-Faber. *Friezen op sosjale media: Rapportazje ûndersyk Taalfitaliteit II* [Frisians on social media: Report research language vitality II], 2015. Fryske Akademy – Mercator European Research Centre on Multilingualism and Language Learning.
- L. Jongbloed-Faber. Frisian on social media: the vitality of minority languages in a multilingual online world, 2021. URL <https://doi.org/10.26481/dis.20210903lj>.
- Lysbeth Jongbloed-Faber, Hans Van de Velde, Cor Van der Meer, and Edwin Klinkenberg. Language use of frisian bilingual teenagers on social media. *Treballs de sociolingüística catalana*, pages 27–54, 2016. URL <https://publicacions.iec.cat/repository/pdf/00000238/00000080.pdf>.
- A. Kilgarriff, S. Reddy, J. Pomikálek, and A. PVS. A corpus factory for many languages. In *In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 2010. Universitat d'Alacant, European Language Resources Association (ELRA). URL [https://www.sketchengine.eu/wp-content/uploads/2015/05/Corpus\\_Factory\\_2010.pdf](https://www.sketchengine.eu/wp-content/uploads/2015/05/Corpus_Factory_2010.pdf).
- H. Kuipers-Zandberg and R. Kircher. The objective and subjective ethnolinguistic vitality of West Frisian: Promotion and perception of a minority language in the Netherlands. *Sustainable Multilingualism*, 17:1–25, 2020. URL <https://doi.org/10.2478/sm-2020-0011>.

- R. Landry and R.Y. Bourhis. Linguistic landscape and ethnolinguistic vitality: An empirical study. *Journal of Language and Social Psychology*, 16(1):23–49, 1997.
- C. Moseley. Atlas of the world's languages in danger (3rd edn.), 2010. URL <http://www.unesco.org/culture/en/endangeredlanguages/atlas>.
- P. Nicolai, M. Swart, S. ter Braak, and B. Doets. 'Krenten uit de pap' - Op zoek naar de meerwaarde van de Friese taal in het economisch verkeer ['Icing on the cake' - In search of the added value of Frisian the course of trade], 2020. E&E Advies.
- Omrop Fryslân. *Mear as heal miljoen oersettings yn Google Translate Wike* [More than half a million translations in Google Translate Week], 2015. URL <https://www.omropfryslan.nl/nijs/575304-mear-heal-miljoen-oersettings-yn-google-translate-wike>. Omrop Fryslân.
- Omrop Fryslân. *Praat (noch) mar Frysk* [Speak Frisian (still)], 2018. URL <https://www.omropfryslan.nl/utstjoering/fryslan-dok-fan-14-oktober-2018-1700>. [documentary]. Omrop Fryslân.
- Omrop Fryslân. *Gjin Fryske tolken mear beskikber yn rjochtbank fanwege 'meagere fergoeding'* [No more Frisian interpreters available in court because of 'meagre compensation'], 2021a. URL <https://www.omropfryslan.nl/nijs/1094557-gjin-fryske-tolken-mear-beskikber-yn-rjochtbank-fanwege-meagere-fergoeding>. Omrop Fryslân.
- Omrop Fryslân. *Mozilla-projekt kin noch wol mear ynsprekkers Frysk brûke* [Mozilla project can use more speakers Frisian], 2021b. URL <https://www.omropfryslan.nl/nijs/1095205-mozilla-projekt-kin-noch-wol-mear-ynsprekkers-frysk-bruke>. Omrop Fryslân.
- Provinsje Fryslân. *Fryske Taalatlas 2020* [Frisian Language Atlas 2020], 2020a. Provinsje Fryslân.
- Provinsje Fryslân. *Nij Poadium:Keunst & kultuer, taal & ûnderwiis en erfgoed yn Fryslân 2021-2024* [New Stage: Art & culture, language & education and heritage in Fryslân 2021-2024], 2020b. URL [https://www.fryslan.frl/\\_flysystem/media/nota-nij-podium-kultuer-taal-en-erfgoed-2021-2024-frl.pdf](https://www.fryslan.frl/_flysystem/media/nota-nij-podium-kultuer-taal-en-erfgoed-2021-2024-frl.pdf). Provinsje Fryslân.
- Georg Rehm and Hans Uszkoreit, editors. *META-NET White Paper Series: Europe's Languages in the Digital Age*, 32 volumes on 31 European languages, Heidelberg etc., 2012. Springer.
- Georg Rehm, Katrin Marheinecke, Stefanie Hegele, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Khalid Choukri, Andrejs Vasiljevs, 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 Auksoriütë, Núria Bel, António Branco, Gerhard Budin, Walter Daelemans, Koenraad De Smedt, Radovan Garabik, 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. The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe. In 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, editors, *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, pages 3315–3325, Marseille, France, 5 2020. European Language Resources Association (ELRA).
- Rijksoverheid. *Welke erkende talen heeft Nederland* [Which recognised languages does the Netherlands have?], n.d. URL <https://www.rijksoverheid.nl/onderwerpen/erkende-talen/de-friese-taal>. Retrieved January 13, 2022.
- A. F. Schukking and E.L. Klinkenberg. *De rol fan it Frysk by gemeentlike weryndielingen yn Fryslân* [The role of Frisian in municipal amalgamations in Fryslân], 2018. URL [https://www.mercator-research.eu/fileadmin/mercator/documents/publications/Weryndieling\\_Fryske\\_Gemeenten\\_-\\_Webmin.pdf](https://www.mercator-research.eu/fileadmin/mercator/documents/publications/Weryndieling_Fryske_Gemeenten_-_Webmin.pdf). Mercator European Research Centre on Multilingualism and Language Learning.

- Sintrum Frysktalige Berne-opfang. *Algemien* [General], n.d. URL <https://www.sfbo.nl/fri/algemien>. SFBO. Retrieved January 13, 2022.
- M. Sloos, E. Drenth, and W. Heeringa. The Boarnsterhim Corpus: A Bilingual Frisian-Dutch Panel and Trend Study. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, 2018. European Language Resources Association (ELRA). URL [https://www.researchgate.net/publication/326209416\\_The\\_Boarnsterhim\\_Corpus\\_A\\_Bilingual\\_Frisian-Dutch\\_Panel\\_and\\_Trend\\_Study](https://www.researchgate.net/publication/326209416_The_Boarnsterhim_Corpus_A_Bilingual_Frisian-Dutch_Panel_and_Trend_Study).
- Peter Spyns and Elisabeth D'Halleweyn. The STEVIN Programme: Result of 5 Years Cross-border HLT for Dutch Policy Preparation. *Essential Speech and Language Technology for Dutch*, pages 21–39, 2013.
- Frieda Steurs, Vincent Vandeghinste, and Walter Daelemans. Deliverable D1.10 Report on the Dutch Language, 2022. URL [https://european-language-equality.eu/wp-content/uploads/2022/03/ELE\\_\\_\\_Deliverable\\_D1\\_10\\_Language\\_Report\\_Dutch\\_.pdf](https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D1_10_Language_Report_Dutch_.pdf). Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.
- Alan M. Turing. Computing machinery and intelligence. *Mind*, LIX(236):433–460, 1950. ISSN 0026-4423. doi: 10.1093/mind/LIX.236.433. URL <https://doi.org/10.1093/mind/LIX.236.433>.
- H. Van de Velde. *It fertutearzjen fan it Frysk* [The neglect of Frisian], 2021. URL [https://www.fryske-akademy.nl/fileadmin/inhoud/beelden/homepage/Nijs\\_en\\_aginda/Eveneminten/HansVandeVelde\\_It\\_fertutearzjen\\_fan\\_it\\_Frysk\\_-\\_De\\_verwaarlozing\\_van\\_het\\_Fries\\_DEFweb\\_.pdf](https://www.fryske-akademy.nl/fileadmin/inhoud/beelden/homepage/Nijs_en_aginda/Eveneminten/HansVandeVelde_It_fertutearzjen_fan_it_Frysk_-_De_verwaarlozing_van_het_Fries_DEFweb_.pdf). Fryske Akademy.
- H. Van den Heuvel, M. Bentum, L. Ten Bosch, and S. Wills. A speech recognizer for subtitling Frisian/Dutch council meetings, 2021. URL [https://www.nvfw.org/sites/nvfw.org/files/Abstract\\_DvdF2021\\_VanDenHeuvel\\_et\\_al.pdf](https://www.nvfw.org/sites/nvfw.org/files/Abstract_DvdF2021_VanDenHeuvel_et_al.pdf). CLST, Radboud University, Nijmegen.s.
- Nynke Varkevisser and Albert Walsweer. It is mei sizzen net te dwaan: Rapport taalplan frysk [it is impossible to say. taalplan frysk report. inventory of the state of affairs regarding (the subject of) frisian in primary and secondary education]. *Ljouwert/Leeuwarden: Provinsje Fryslan*, 2018.
- Wettenbank. Wet gebruik Friese taal [Use of the Frisian Language Act], 2022a. URL <https://wetten.overheid.nl/BWBR0034047/2014-01-01>. Retrieved March 23, 2022.
- Wettenbank. Tijdelijk besluit digitale toegankelijkheid overheid [Temporary decision on the digital accessibility of the government], 2022b. URL <https://wetten.overheid.nl/BWBR0040936/2018-07-01>. Retrieved March 23, 2022.
- E. Yilmaz, J. Dijkstra, H. Van de Velde, F. Kampstra, J. Algra, H. van den Heuvel, D. van Leeuwen, S. Kingma, and M. Andringa. A Longitudinal Bilingual Frisian-Dutch Radio Broadcast Database Designed for Code-switching Research. 2016. URL [https://pure.knaw.nl/admin/files/5834638/LREC2016\\_1\\_P.pdf](https://pure.knaw.nl/admin/files/5834638/LREC2016_1_P.pdf).