MULTILINGUAL EUROPE: FACTS, CHALLENGES, OPPORTUNITIES

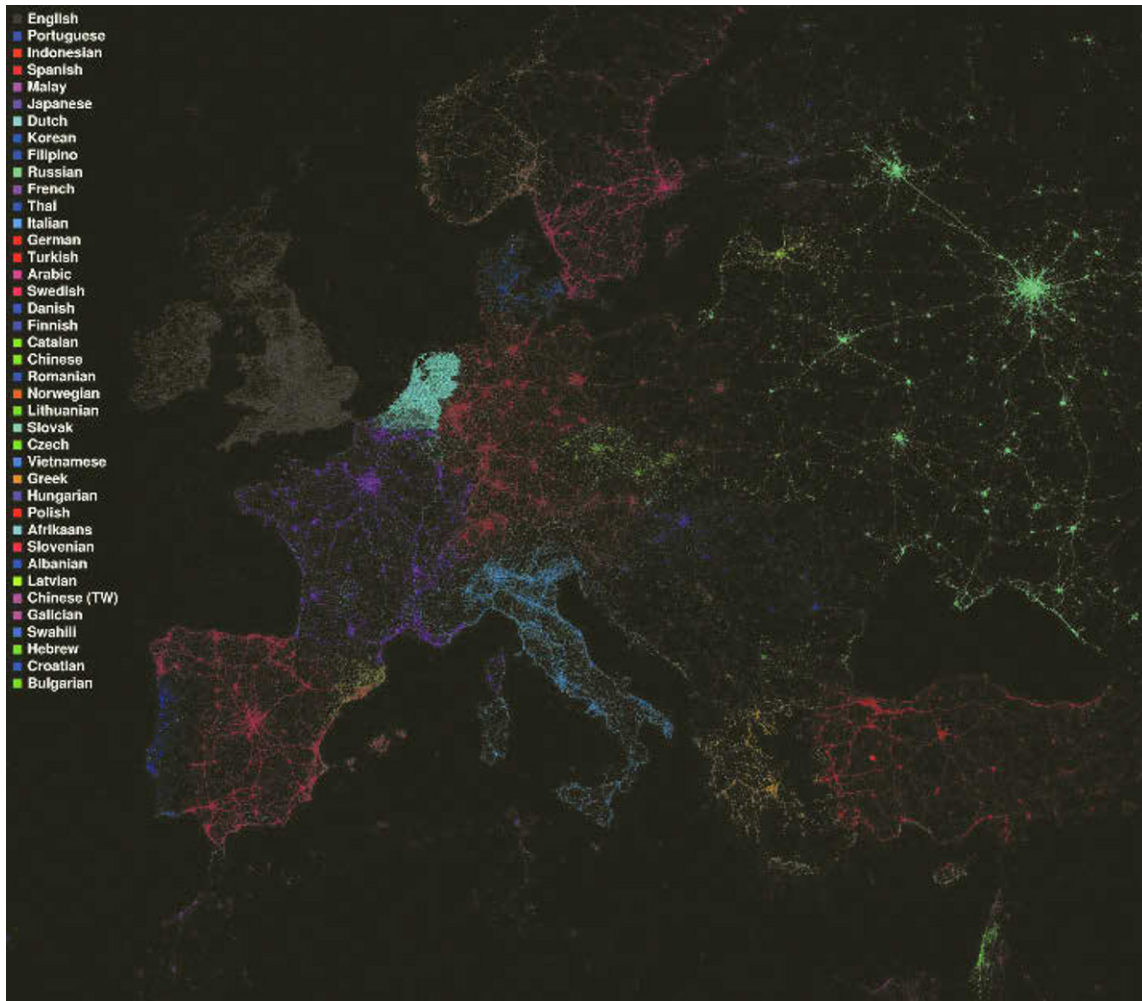## 2.1 EUROPE'S LANGUAGES IN THE NETWORKED SOCIETY

Europe's more than 80 languages are one of its richest and most important cultural assets, and a vital part of its unique social model [6, 11]. While languages such as English and Spanish are likely to thrive in the emerging digital marketplace, many European languages could become marginal in a networked society. This would weaken Europe's global standing, and run counter to the goal of ensuring equal participation for every European citizen regardless of language. A recent UNESCO report on multilingualism states that languages are an essential medium for the enjoyment of fundamental rights, such as political expression, education and participation in society [18, 19, 20, 21]. From the very beginning, Europe had decided to keep its cultural and linguistic richness and diversity alive during the process of becoming an economic and political union. For maintaining the policy of multilingualism, the EU's institutions spend about one billion Euros a year on translating texts and interpreting spoken communication. For all European economies the translation costs for compliance with the laws and regulations are much higher.

A single European market that secures wealth and social well-being is possible, but linguistic barriers still severely limit the free flow of goods, information, services, debates and innovation. With the increased number of EU members and the general trend towards timely trans-border interaction, everyday communication between Europe's citizens, within business and among politicians is more and more becoming confronted with language barriers. Many Europeans find it difficult to interact with online services and participate in the digital economy. According to a recent study, 57% of internet users in Europe purchase goods and services in languages that are not their native language (English is the most common foreign language followed by French, German and Spanish). 55% of users read content in a foreign language while only 35% use another language to write e-mails or post comments on the web [22]. A few years ago, English might have been the lingua franca of the web – the vast majority of content on the web was in English – but the situation has now drastically changed. The amount of online content in other European as well as Asian and Middle Eastern languages has exploded [23]. Already today, more than 55% of web-based content is not in English. One language is especially becoming more and more dominant: a recent study by the UN Broadband Commission reports that Chinese internet users will overtake English language users by 2015 [24].

Figure 3 shows the European language communities of Twitter: the map was created by identifying automatically the languages millions of tweets are written in and placing them onto a map using their GPS-coordinates [25]. To a large degree the resulting map replicates Europe's language borders – and barriers.

Surprisingly, this ubiquitous digital divide due to language borders and language barriers has not gained much

English
Portuguese
Indonesian
Spanish
Malay
Japanese
Dutch
Korean
Filipino
Russian
French
Thai
Italian
German
Turkish
Arabic
Swedish
Danish
Finnish
Catalan
Chinese
Romanian
Norwegian
Lithuanian
Slovak
Czech
Vietnamese
Greek
Hungarian
Polish
Afrikaans
Slovenian
Albanian
Latvian
Chinese (TW)
Galician
Swahili
Hebrew
Croatian
Bulgarian

3: Language communities of Twitter (European detail) [25]

public attention up until our recent press campaign in which we informed the public about the findings of our META-NET study "Europe's Languages in the Digital Age" (see Chapter 4, p. 27 ff.). In this study, published in our META-NET White Paper Series [12], more than 200 experts from all over Europe found that at least 21 of the 30 languages examined are in serious danger of facing digital extinction. A pressing question raises: which European languages will thrive in the networked information society, and which are doomed to disappear?

The European market for translation, interpretation and localisation was estimated to be 5.7 billion Euros in 2008. The subtitling and dubbing sector was at 633 million Eu-

ros, language teaching at 1.6 billion Euros. The overall value of the European language industry was estimated at 8.4 billion Euros and expected to grow by 10% per year, i. e., resulting in ca. 16.5 billion Euros in 2015 [26, 27]. (The global speech technology market is even bigger, it will reach ca. 20.9 billion US-Dollars by 2015 and ca. 31.3 billion US-Dollars by 2017 [28].) Yet, this existing capacity is not enough to satisfy current and future needs, e. g., with regard to translation [29]. Already today, Google Translate translates the same volume per day that all human translators on the planet translate in one year [30].

Despite recent improvements, the quality, usability and integration of machine translation into other online ser-

vices is far from what is needed. If we rely on existing technologies, automated translation and the ability to process a variety of content in a variety of languages – a key requirement for the future internet – will be impossible. The same applies to information services, document services, media industries, digital archives and language teaching. There is an urgent need for innovative technologies that help save costs while offering faster and better language services to the European citizen.

The most compelling solution for ensuring the breadth and depth of language usage in tomorrow's Europe is to use appropriate technology. Still, despite recent improvements, the quality and usability of current technologies is far from what is needed. The META-NET study mentioned above shows that, already today, especially the smaller European languages suffer severely from underrepresentation in the digital realm. There are tremendous deficits in technology support and significant research gaps for all languages. For example, machine translation support for 23 out of the studied 30 languages was evaluated as having very limited quality and performance, which is an alarming result!

Another important aspect related to the European discourse. Especially the one on innovation has become determined by the English language, and the media reporting in that language. As Mark Vanderbeeken, a Belgian who lives in Italy noted in a widely read essay [31], this sheer dominance of English carries with it an accompanying perspective of Europe, both in terms of stereotypes and in terms of relevance to the Anglo-Saxon world. This puts European businesses and countries at a serious disadvantage that they are not even aware of. It also disadvantages businesses in the English-speaking world, which are perhaps not aware that they are receiving an abbreviated picture of innovation in Europe. Vanderbeeken calls this phenomenon "the non-English disadvantage". It is another example of a disadvantage which can be successfully addressed through multilingual technologies.

2.2 HOW CAN LANGUAGE TECHNOLOGY HELP?

One way to overcome language barriers is to learn foreign languages. Yet without technological support, mastering the EU's 23 official languages and some 60 other European languages is an insurmountable obstacle for Europe's citizens, economy, scientific progress, and political debate [32]. The solution is to build key enabling technologies: language technologies will offer all European stakeholders tremendous advantages and benefits, not only in the single market, but also in trade relations with non-European countries.

Language technology is also a key enabler for the knowledge society. It supports humans in everyday tasks, such as writing e-mails, searching for information online or booking a flight. It is often used behind the scenes of other software applications. We benefit when we use spelling checkers, browse recommendations in an online shop, hear the spoken instructions of a navigation system or translate web pages with an online service.

Several popular language technology services are provided by US companies, some of them free of charge. The recent success of Watson, an IBM computer system that won against human candidates in the game show Jeopardy, illustrates the immense potential. As Europeans, we urgently have to ask ourselves a few crucial questions:

- Can we afford our information, communication and knowledge infrastructure to be highly dependent upon monopolistic services provided by US companies (technological lock-in)?

- What is Europe's fallback plan in case the language-related services provided by US companies that we rely upon are suddenly switched off or if serious access or security issues arise?

- Are we actively making an effort to compete in the global landscape for research and development in language technology?

- Can we expect third parties from other continents to solve our translation and knowledge management problems in a way that suits our specific communicative, societal and cultural needs?
- Can the European cultural background help shape the knowledge society by offering better, more secure, more precise, more innovative and more robust high-quality language technology?

We believe that *Language Technology made in Europe for Europe* will significantly contribute to future European cross-border and cross-language communication, economic growth [8] and social stability while establishing for Europe a worldwide, leading position in technology innovation, securing Europe's future as a world-wide trader and exporter of goods, services and information.

2.3 SOCIETAL CHALLENGES

Information technology is bringing people speaking different languages together in new ways. Highly popular social networks and social media such as Wikipedia, Facebook, Twitter, YouTube, Google+, Pinterest, and Instagram are only the tip of the iceberg.

Many societal changes and economic as well as technological trends confirm the urgent need to include sophisticated language technology in our European ICT infrastructure. Research, development and innovation efforts in LT must increase to go beyond what is possible today.

Language Barriers. A study on online commerce shows that language barriers are economic barriers [33]. Only 59% of retailers can handle transactions in more than one language. Translation and localisation costs must be drastically lowered before Europe's single digital market is a reality. Multilingual language technology is the key, especially for SMEs. At the same time, 81% of all internet users think that websites run in their country should also be available in other languages. 44% of European users think they miss out on interesting information because

websites are not available in a language they understand [22]. These facts can no longer be ignored. Reliable LT can help establish a vast market for information as well as consumer and entertainment goods in any language.

Ageing Population. Demographic changes bring about a need for more assistive technologies, especially for spoken language access. An ageing population requires technology that can help master everyday situations, provide proactive guidance and that could answer the question, "Where did I leave my glasses?" Also, more health care services and support systems will be required. Ambient assisted living (AAL) technologies can greatly benefit from a personalised, spoken method of interaction.

People with Disabilities. New technologies can help us reach the ambitious goal of achieving equal opportunities and promoting independent living. Language technologies already help people with disabilities to participate in society. Noteworthy examples include screen readers, dictation systems and voice-activated services. In addition to the social aspect there is a huge commercial market for future technologies such as, for example, full dialogue systems and interactive assistants, sign language recognition and synthesis, automatic translation, summarisation and content simplification. Approximately 10% of Europeans (50 million citizens) have permanent disabilities.

Immigration and Integration. According to the United Nations' International Migration Report 2002, 56 million migrants lived in Europe in 2000 [34]. This number has grown to ca. 60 million people today. Facilitating communication, providing access to information in foreign languages and helping people learn European languages can help better integrate migrants into European society. In fact, speech and language technologies can dramatically improve the integration process by providing intelligent language learning environments, automatic subtitling and translation services in real time.

Personal Information Services and Customer Care. In our 24/7 "always on" economy we expect quick and reli-

able answers as well as engaging and timely online news broadcasts. However, information overload still poses a serious problem. Citizens, governments and industries would greatly benefit from new technologies that help get the situation under control again. Language-enabled mobile applications will become personal assistants to everyone, offering automatic and intelligent question answering and dialogue capabilities, as well as automatic, personalised and trusted text and speech processing of messages, news items and other content.

Global Cooperation and Human Communication. Companies need to address new markets where multiple languages are spoken and support multinational teams at multiple locations. Many jobs cannot be filled today because linguistic barriers exclude otherwise qualified personnel. Improvements in language technology can enable richer interactions and provide more advanced tele and video conferencing services. Future technologies like a 3D internet can enable new modes of collaboration as well as support more realistic training and education scenarios. We will soon be able to participate in virtual events as new forms of entertainment, cultural exchange and tourism. Combining virtual worlds and simulations with multilingual language technology including translation, automatic minute taking, video indexing and searching will let us experience being European in a new way.

Preservation of Cultural Heritage and Linguistic Diversity. According to the principles of the UN-endorsed World Summit on the Information Society [35], the "Information Society should be founded on and stimulate respect for cultural identity, cultural and linguistic diversity." Much effort has been put into digital archives such as Europeana that help promote our cultural heritage. However, digitisation is only the first step. The sheer amount of information and language barriers hinder access of our cultural treasures. Language technology can make this content accessible, e. g., through cross-lingual search and machine translation. Likewise, com-

munication skills need to be trained. This is underlined by the UNESCO Information for All Programme, which seeks to "foster the availability of indigenous knowledge through basic literacy and ICT literacy training" [36].

Social Media and e-Participation. Social networks have a significant impact on all areas of society and life. They can help us solve technical problems, research products, learn about interesting places or discover new recipes. Recent developments in North Africa demonstrate their ability to bring citizens together to express political power. Social media will play a key role in the discussion of important, future topics for Europe like a common energy strategy and foreign policy. However, certain groups are becoming detached from these developments. One can even speak of a broken link regarding communication cultures. This is an issue since bottom-up movements are highly relevant for politicians, marketing experts, and journalists who would like to know what customers or citizens think about initiatives, products, or publications. However, it is not possible to process manually the sheer amount of information generated in multiple languages on social networks. We need language technologies that are able to analyse these streams in real time.

Market Awareness and Customer Acceptance. Language technology is a key part of business and consumer software but often hidden inside other, more visible products. Customer acceptance of LT has recently been shown to be high. For example, market research by the Ford Motor Company indicates that their voice control system, Ford SYNC, is widely accepted [37]. 60% of Ford vehicle owners use voice commands in their cars. Non-Ford owners report a three-fold increase in their willingness to consider Ford models while 32% of existing customers admit that the technology played an important role in their purchase decision. Language technology has a tremendous market potential.

One Market, Many Languages. Support for the 23 official languages of the EU has major economic, so-

cial and political implications. Europe currently lags behind countries such as India (22 official languages) and South Africa (11 national languages). Government programmes in these two countries actively foster the development of language technology for a significant number of official languages [38, 39]. Mobile devices are an even more important bridge between humans and information technology. Google already provides free translation services in 3,306 different language pairs as well as voice input for 16 languages and speech output for 24 languages. Apple's App Store has demonstrated how premium content and products can be marketed for free and for a fee. Europe must address this global competition.

Secure Europe. The effective persecution of illegal online activities such as fraud and identity theft requires automatic tools that can help detect crimes and monitor offenders. Language technology can help to build systems that can monitor, analyse and summarise large amounts of text, audio and video data in different languages and from different sources.

This collection of solutions was influenced by bigger trends (see Chapter 3.1). Many of these products and services are only available online. For example, Facebook and Twitter enabled recent political developments in North Africa. In Europe, the idea of social innovation has recently sparked an interest as it "offers an effective approach to respond to social challenges by mobilising people's creativity to develop solutions and make a better use of scarce resources" [40]. Social innovation is part of Europe's 2020 strategy and critically relies on active involvement of citizens, which in turn calls for supportive multilingual language technologies.

Multilingualism has become the global norm rather than the exception [19]. Future applications that embed information and communication technology require sophisticated language technologies. Fully speech-enabled autonomous robots could help in disaster areas by rescuing travellers trapped in vehicles or by giving first aid. Lan-

guage technology can significantly contribute towards improving social inclusion. Language technology can help us provide answers to urgent social challenges while creating genuine business opportunities. Language technology can now automate the very processes of translation, content production, and knowledge management for all European languages. It can also empower intuitive language/speech-based interfaces for household electronics, machinery, vehicles, computers and robots.

In addition to these vertical societal challenges there are multiple horizontal properties that future language technologies need to exhibit. One of these properties is situation or context awareness. Many or even most applications sketched above need to exhibit a certain level of situation or context awareness. The challenge is to design and implement a paradigm in which language technologies are no longer static applications but able to adapt themselves to specific situations, trends and contexts such as, for example, user preferences or user interests. Security applications need to be aware of criminal or violent tendencies in communication patterns. E-participation systems need to be aware of interest in societal issues and need to have access to internet debates and the opinion of large online communities towards certain topics. Tools for the analysis of market awareness need methods for reputation mining, customer relationship systems need algorithms for attitude analysis.

2.4 MARKET OPPORTUNITIES

The market offers tremendous business potential for European language technology companies, especially for online retailers, language services, LT usage in key markets, data intensive scenarios and selected devices and environments. (This section is partially based on [41]).

Most online retailers are limited to small segments, the largest of which scarcely exceeds 60 million in population [42]: 82% of European retailers operate in only a single language, 11% in only two, and only 2% provide services

in five or more languages; only 21% of European retailers support cross-border transactions. Although 51% of European retailers sell via the internet, a vanishing small number of Europeans currently engage in online cross-border purchases. Language technology that lowers the burden and costs of translation and localisation for European languages would not only open the European market to European businesses, but enable them to access the estimated population of one billion individuals worldwide who speak one of Europe's major languages, with accompanying economic benefit for European companies. The market for LT software is currently expected to grow to 30 billion Euros by 2015 (versus 20 billion Euros today). European enterprises – particularly the more than 500 active European SMEs – have the potential to dominate the field if they can offer compelling solutions to Europe's needs for online businesses and other fields.

Aside from the sales potential for online retailers, deployment of LT would increase overall demand for language-related services, currently worth ca. 5 billion Euros in Europe (expected to grow to ca. 8 billion Euros by 2015). As translation becomes the norm rather than the exception, the translation market, one which Europe currently dominates, would be expected to see substantially faster growth than anticipated.

In addition, dedicated LT-intense services will gain importance. Examples are technical translation supported by LT in the automotive domain, automatic interpreting for tourism and culture, or speaker verification for financial services and banking. The European LT industry is in a good position to serve these markets, since European LT companies specialise in these domains.

The business role of LT can be characterised in terms of its relation to the Big Data market (estimated at ca. 4 billion Euros in 2012, expected to grow to ca. 13 billion Euros by 2015), cloud-based models for distribution and computation (expected to reach 45 billion Euros in the near future) or business data intelligence gathering and analysis (currently a 27 billion Euros market). In all these areas LT will be crucial for assuring high quality and meaningful use of data and data infrastructures.

Finally, certain types of systems and devices will require LT for core functionality. Mobile devices currently drive 43% of current IT growth; embedded systems are currently an 800 billion Euros p. a. industry. At the moment U.S.-based companies have a lead in these areas, but their offerings often do not consider multilingualism as a base requirement. This will create a market opportunity in the billions of euros for European LT companies.