3

MAJOR TRENDS IN INFORMATION AND COMMUNICATION TECHNOLOGIES

3.1 THE CURRENT STATE

Networked computers are ubiquitous. They come in different shapes and forms (desktop, laptop, mobile phones, tablets, ebook readers, etc.) or are embedded in devices, objects, and systems such as, for example, cameras, washing machines, televisions, cars, heating systems, robots, traffic control systems. Software is usually available in multiple human languages. Global standardisation efforts such as Unicode solved the problem of representing and displaying different alphabets and special characters.

Mobile devices and social media are reshaping how and when we communicate with one another using the tools and devices we use both in business and private life. The way we interact with computers is no longer restricted to graphical interfaces and keyboards, but it is being extended through touch screens, voice interfaces and dialogue systems, and mobile devices with accelerometers that tell the device how it is held by the user.

Language technology is currently not well integrated into applications and interfaces – to the end user, spelling, grammar checking and maybe search seem to be the only notable exceptions. Apple's introduction of the mobile assistant Siri on the iPhone and a similar product by Google illustrate the trend towards more intelligent language-based interaction.

The web represents much of our knowledge. It emerged as a collection of static documents. Nowadays it is first and foremost a collection of systems and databases that can be queried through APIs, and applications such as Google Mail, Facebook, eBay and Amazon. Many people only need one application on their computers: a web browser. Others use netbooks whose operating system more or less *is* the browser (Chromium OS). Behind the scenes, there is already a considerable amount of language technology incorporated in web applications such as search engines, dialogue systems, or machine translation services but these are not immediately visible or recognisable by the user as language technologies as such.

3.2 HARDWARE AND SOFTWARE

Networked computers come in many shapes and forms, from mobile phones to tablets, netbooks, ultra-portable laptops, small desktop computers and ebook readers to devices such as radios, televisions, gaming consoles and other entertainment devices with built-in wireless and access to, for example, RSS feeds, internet radio stations or youtube, cameras or house-hold appliances such as fridges, coffee machines or scales that push the user's weight to the cloud from where it can be monitored using an app on the smartphone. The next hardware revolution will be wearable computers. Google has already demonstrated a prototype of their Google Glasses product in which the computer visuals are projected into a head-up display. This approach can be used to provide the user with a true augmented reality perspective and a hands-free computing environment which immediately brings

G. Rehm and H. Uszkoreit (eds.), *META-NET Strategic Research Agenda for Multilingual Europe 2020*,
White Paper Series, DOI: 10.1007/978-3-642-36349-8_3, © The Author(s) 2013

up the question how to interact with this device – by using only your voice?

The shape and size of computers is no longer determined by the shape and size of their internal hardware components. Due to breakthroughs in miniaturisation, their form now truly follows their function. While computers and devices with embedded systems get smaller and smaller, the distributed data centres around the world get bigger and bigger – both in terms of number and size. The concept of cloud computing and storing data in dedicated data centres from where the data can be accessed by multiple devices, is already mainstream and used by millions of consumers world-wide. An important reason for the cloud's success is the fact that, by now, people tend to have more than one computer. A not too unusual setup may include a laptop, a smartphone, a tablet and another computer as a dedicated media centre. Cloud services are ideal for synchronising data between many devices.

The trends in the software area are much more multi-dimensional. Here we can only scratch the surface and highlight several recent developments and current trends.

Communication: A cornerstone of today's computer use is communication, be it more direct communication via traditional e-mail, instant messaging, text-based chat systems, video chat between two people or larger groups or indirect communication and staying in touch with friends, acquaintances and colleagues via social networks such as Twitter, Facebook, LinkedIn, Instagram or social media such as blogs, YouTube, or Pinterest. Millions of people world-wide are always online using several different networked devices including their phones.

Search and Information Services: An important use case of any type of device is to search for information and to make use of information services. Important applications are web search engines, online encyclopedias, news sites, digital libraries such as Europeana, meta-search engines and RSS feed aggregators etc.

Location-based Services: Search queries are often coupled to the user's location. Location-based services enable the user to search for information in his or her geographic area, to make use of online maps, navigation systems, recommender systems or to find tweets or photos from the neighbourhood.

Media monitoring: Search and retrieval enable users to find information they already know about or suspect exist. Both are about finding the needles in the haystack. Media monitoring and applications with a certain level of situation awareness are not about finding documents or items, they are about keeping track of the state of the world. Applications for this purpose are coming to the market at a rapid pace.

E-Commerce and Shopping: World-wide billions of Euros are spent each year using general online shops such as Amazon or eBay or shops run by specific brands or services, reservation and booking, online banking and brokering services etc.

Media and Entertainment: Different types of media (photos, videos, music, sounds, text and multimedia documents, audio and video podcasts, ebooks, films, tv programmes etc.) play an important role. Not only personal media such as photos or videos and other user-generated content are often posted to social networks, songs, photos or videos created by third parties are also often shared using social networks. Almost all of the media mentioned above can be purchased in online stores, for consumption on any device. Another important software category is games, from online Flash games to games that are embedded into social networks, location-based games, multiplayer games with millions of users to very simple but also very successful casual games such as Angry Birds.

App and Media Stores: The success of ecommerce platforms [17], online shopping and the increased use of digital media led to app and media stores. By now it is possible to buy or rent almost every movie ever made, to buy music, to stream music from the cloud onto your device

and to buy software and mobile apps through dedicated stores without any need to ship physical media. An important development is in-app purchasing, especially on mobile devices: with a single tap of a finger it is possible to buy, within a specific app, additional modules, components or data sets for a small price.

Personal Information Management: With the ever increasing number of personal and professional contacts (including social networks), meetings and personal errands to run, there is a big trend towards personal information management. This includes address and contacts databases that are often integrated into larger applications such as Google Contacts (embedded in, among others, Google Mail) or Apple's AddressBook (used in Apple Mail). Cloud-integration is an important feature, so that contact information (including names, email addresses, phone numbers, photos etc.), calendar entries and "to do" items are always available on all devices.

Office Applications: The classic office applications – word processors, spreadsheets, presentations – are still important in the professional context and also in home use. Nowadays, there are several applications to choose from including open source software, cloud-based services and applications for Apple's iOS. Almost all office suites use the cloud to enable the user to, for example, finish work on a presentation at the desktop computer where the document is automatically pushed to the cloud and to continue working on the presentation on a mobile device on the way home.

One of the most basic common denominators of all pieces of software is language which plays a central and integral part in practically every single tool or application. However, language technology as such (including text analysis, information retrieval and extraction, spelling and grammar checking, speech recognition and synthesis, dialogue systems etc.) is usually completely hidden, integrated into bigger applications, working behind the scenes. There is, however, a clear trend to embed language technologies not only at the level of the single application but on the level of the operating system. Another important factor of current computing is communicating and interacting with other people or groups of people, both on the personal level and also for business purposes. A third crucial ingredient of computing today is information, especially structured information which is annotated based on specific standards (see, for example, the family of standards around XML, Semantic Web, Linked Open Data, Web Services, Big Data etc.).

3.3 CURRENT TRENDS AND MEGA-TRENDS

In the following we sketch some of the current trends and mega-trends, grouped into three sections.

Internet: The internet will continue to be *the* main driving force behind future developments in information and communication technologies. There are several mega-trends tightly coupled to the internet and network technologies: among these are cloud computing and cloud services, including cloud storage, as well as linked open data and the semantic web. Social media and social networks will continue to change everything and to penetrate the market further, including niche markets, driven by location-based services. With the predominance of social networks we expect a certain convergence of digital identities that will enable users to have and to maintain one central digital identity that feeds into their multiple social network profiles. Exchanging and distributing personal data and information (photos, videos, music etc.) in a secure way will become easier. We further expect more broad deployment and general acceptance of services in the areas of e-democracy and e-government (including open data portals) and a continued increase of e-commerce platforms [17] and services. A perceived general information overload will continue to be a problem, although modern search engines, aggregation ser-

vices and user interfaces help a lot. New business models and ways to distribute content or services to the end-user will continue to emerge (see the different app stores and approaches such as in-app purchases).

People: Information and communication technologies are used by people – the predominance of social networks and being always-on using smartphones, tablets and laptops, is responsible for the fact that the way people interact, communicate and do business with one another will continue to be redefined and reshaped completely, including novel approaches for participation and public deliberation processes. Communication tools such as email, twitter, facebook etc. are mainstream by now and used across all age groups. This trend will continue. The trend to use social networks and location-based services to find "faces and places", items or places of interest or new acquaintances with similar hobbies will continue (along with a more in-depth discussion of privacy issues). We expect a tighter connection between the data stored in social networks as well as tools for personal information management and the linked open data cloud.

Hardware and Software: Many internet companies operate under the slogan "mobile first". Accessing the web on mobile devices will overtake the use of desktops and laptops very soon. There is also a tendency for completely novel mobile devices with Apple's iPad and Google's Glasses being two prime examples. More and more household-appliances get connected to the internet (tv, radio, gaming consoles, refrigerator, scales, coffee machine, lamps etc.), ultimately leading to the Internet of Things. Many of these devices will not have displays but voice-driven interfaces. We expect a seamless integration of mobile devices into the hardware landscape at home including simplified data and application transfer and exchange among arbitrary mobile or stationary devices, playing music or movies on displays or video projectors etc. Very soon there will not be a need anymore for the average user to own a laptop or desktop computer because mobile devices will cover all basic needs. The capacity and bandwidth of networks will continue to grow, mobile telecommunication networks will gradually become more important than, for example, ADSL lines. The quality of voice or video calls will continue to improve, phones and all other devices will continue to become faster, have more storage as well as 3D-capable displays that offer more intricate modes of interaction. Mobile phones will have built-in facilities to replace credit cards for payment purposes, effectively replacing the wallet. Finally, the market for apps, especially mobile apps, will continue to grow. Nowadays many companies, services and events have their own app that users can interact with and that usually offer added value when compared to the respective website. Usability will continue to be a decisive factor: only those apps will be successful that users can interact with intuitively right away.

Information and communication technologies will continue to be ubiquitous, available wherever and whenever needed. They will combine widely distributed applications, resources and data and will be able to adapt to the location, situation and needs of the user including current emotions, habits and goals. As can be seen by the success of Wikipedia and other collaboratively edited knowledge bases, it is only a matter of time until one or more gigantic digital models of our world will exist that consist of interlinked and overlapping components. Naturally, languages and especially the automatic processing of languages using language technologies will play a key role in this development. Now is the time to realise the needed breakthroughs. High performance, robust machine translation and related language technology services are urgently needed. There is a huge window of opportunity for consumer-oriented language technology.

Large global platforms for end-user-services have become the predominant innovation drivers for language technology solutions. Well known examples are web services such as Google Search, now integrating the new Knowl-

edge Graph concept network, speech-enabled search [43], web translation services, social networks such as Twitter and Facebook, and combinations of hardware and operating systems such as iOS or Android. The trend towards widely used platforms will drastically facilitate the spreading of innovative language technologies. LT has a good chance of becoming *the* essential feature for the success of the next generation of platforms and services. At closer inspection, the integration of LT in current platforms is very limited, scratching only the surface of what will be possible in the near future.

3.4 SELECTED TREND: BIG DATA, LINKED OPEN DATA AND THE DATA CHALLENGE

There are two important trends concerning data on the web. First, the web is becoming translingual, with content and knowledge being accessible across languages, allowing users to search for and interact with knowledge, but also with devices which are part of the Web of Things, accessible for everybody in their own language. Second, more and more amounts of data – *Big Data* – are being made available online. Big data leads to new challenges in terms of scalability, but also to many new innovations and application scenarios.

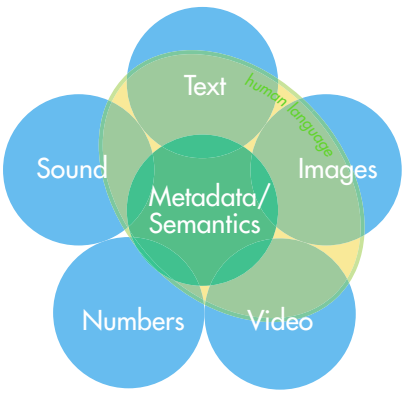
The Translingual Web will enable world wide, borderless communication and commerce. Linked Open Data based on the Semantic Web will be able to support language technologies for improved quality, e. g., in machine translation or cross-lingual search. On the other hand, language technology can support Linked Open Data. It provides the means to create inter- and intra language links and relations to textual knowledge.

Our three priority themes (see Chapter 6) are related to the Translingual Web and data. The Translingual Cloud will benefit from data available across languages. Translation technologies will also help to address data challenges, like building and cleaning data sets that span across languages or providing links between data sets within one or between languages. Multilingual access is an important requirement for a European vision of e-government and e-participation services. On the one hand, language technology can make use of open, governmental data that is made available on portals such as data.gov.uk or within the upcoming European data portal. On the other hand, improving language technologies is inevitable for realising multilingual access to public sector data for all European citizens, as recommended by the European Interoperability Framework for European public services [44]: the sheer amount of data and language barriers between data sets are obstacles that can only be removed with language technologies (e. g., machine translation, cross-lingual information access and information extraction). Finally, one application scenario of Socially-Aware Interactive Assistants are multilingual virtual meetings that make use of shared data sets that provide information about individuals, organisations and interaction settings.

In order to be able to overcome language barriers, data infrastructures need to be made available, while carefully taking licensing and data provenance into account. Existing language and localisation resources (e. g., terminological, lexical data or translation memories) need to be transformed into linked open data. Only then will they be able to play a key role for creating truly multilingual linked open data. Standardisation is crucial when it comes to implementing the infrastructure. So are reference implementations that deal with standardised data and metadata for human language in LT, localisation, CMS, CAT and TMS tools, to assure that standards can be put into action easily and get wide adoption.

Language Technology will also play a key role for Big Data (Figure 4). Building future-proof solutions for big data analysis is impossible without Language Technology. Big data analysis will not be slightly better if we include language technology – it simply will not happen. We cannot

download big data into a database and then build applications on top of it – we will need to process it sensibly and that sense will need to be based on language. This challenge not only relates to structured big data but also to any type of unstructured data including text documents and social media streams, esentially any sequential symbolic process of meaningful information. LT will build bridges from big data to knowledge, from unstructured data to structured data, and can finally lead to what some people are already referring to as Big Semantics. Language Technology will become the foundation for organising, analysing and extracting data in a truly useful way.



4: Human language in the world of data

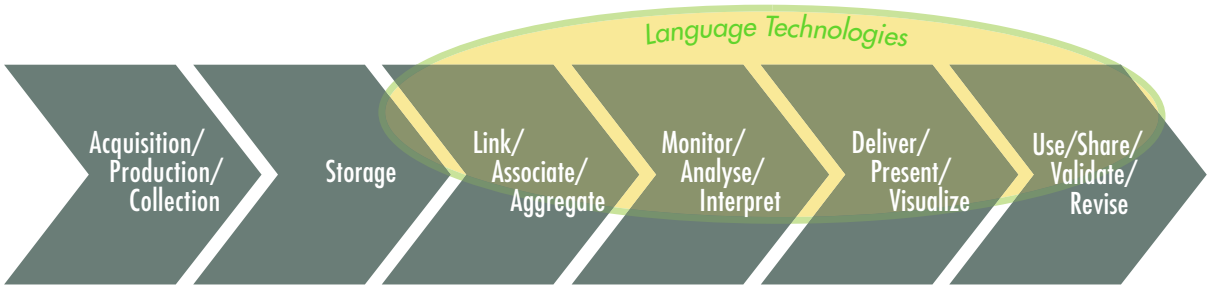
To achieve success in these trends, various prerequisites need to be fulfilled. Linked open data sets need to be enriched with multilingual information. For textual knowledge we can expect that the enrichment will trigger a bootstrapping process. Here, bootstrapping means that existing Semantic Web vocabularies and data sets will be enriched with multilingual information in a first step. They can then be exploited as background knowledge for improved text analysis. Afterwards they can be fed back into the world of linked open data. Models such as, e. g., Lemon for enriching ontologies with multilingual, linguistic information will lead to richer resources and quality in the areas of machine translation, question answering, information extraction or textual entailment. This will create a synergetic cycle, in which the Semantic Web and deep text analysis benefit from each other, effectively bootstrapping the Translingual Web. For realising these synergies, methodologies need to be developed both for high quality, manually created linked open data resources and for big data, e. g., analysing activities of billions of users on the global, multilingual social web.

Another pre-requisite for the convergence of data and LT is the availability of free, open and interoperable data sources. Existing resources such as Wikipedia, DBpedia, Wikidata, Yago and OpenStreetMap need to be consolidated, based on standardised vocabularies to support interoperability and re-use. Core ontology vocabularies need to be translated into different languages. We need tools for cleaning up data, as well as mechanisms that can aggregate, summarise and repurpose content. For all LT applications that interact with data, the regulation of intellectual property rights is a problem that needs to be resolved soon. The web is a global space, and Europe has to find a legal approach that supports both local R&D while fostering global competitiveness.

In FP7, projects and efforts such as DBpedia, Monnet, Wikidata and META-NET's META-SHARE have started tackling some of the problems discussed above. Organisations like ISO TC37 SC4, GALA and the World Wide Web Consortium support this work by providing standardised building blocks for application development and data sharing. Europe is in a good position to be in the driver's seat of the data challenge, both for human knowledge and big data, effectively creating multiple new data value chains (Figure 5).

3.5 SELECTED TREND: FROM CLOUD COMPUTING TO SKY COMPUTING

A major megatrend is known as cloud computing. A large proportion of IT solutions is already offered through the internet, forecasts predict that it will increase rapidly.



5: Language technology in the Data Value Chain

Computing may be offered on different levels of abstraction ranging from Infrastructures as a Service (IaaS) via Platforms as a Service (PaaS) to the powerful concept of providing any suitable software product as an internet service (Software as a Service, SaaS). Especially the latter concept has far-reaching, mainly beneficial, implications for distribution, support, customisation, maintenance and pricing. It also opens new opportunities for software evolution by emerging dynamic schemes of integration, evaluation, adaptation and scaling. A well-known example are the Google Docs office applications. In language technology an increasing number of solutions are already offered as free or commercial web services, among them machine translation, language checking and text-to-speech conversion.

A special challenge for cloud computing is the need for trust. Since the services are rendered outside their sphere of control, customers demand sufficient safeguards securing performance, data protection, and persistence. Large European users of translation technology do not send their corporate language data to the existing large online translation services because the service providers do not offer a trust mechanisms. The situation is even more severe for business intelligence applications where the confidentiality of the collected information can be mission critical for planning and decision processes.

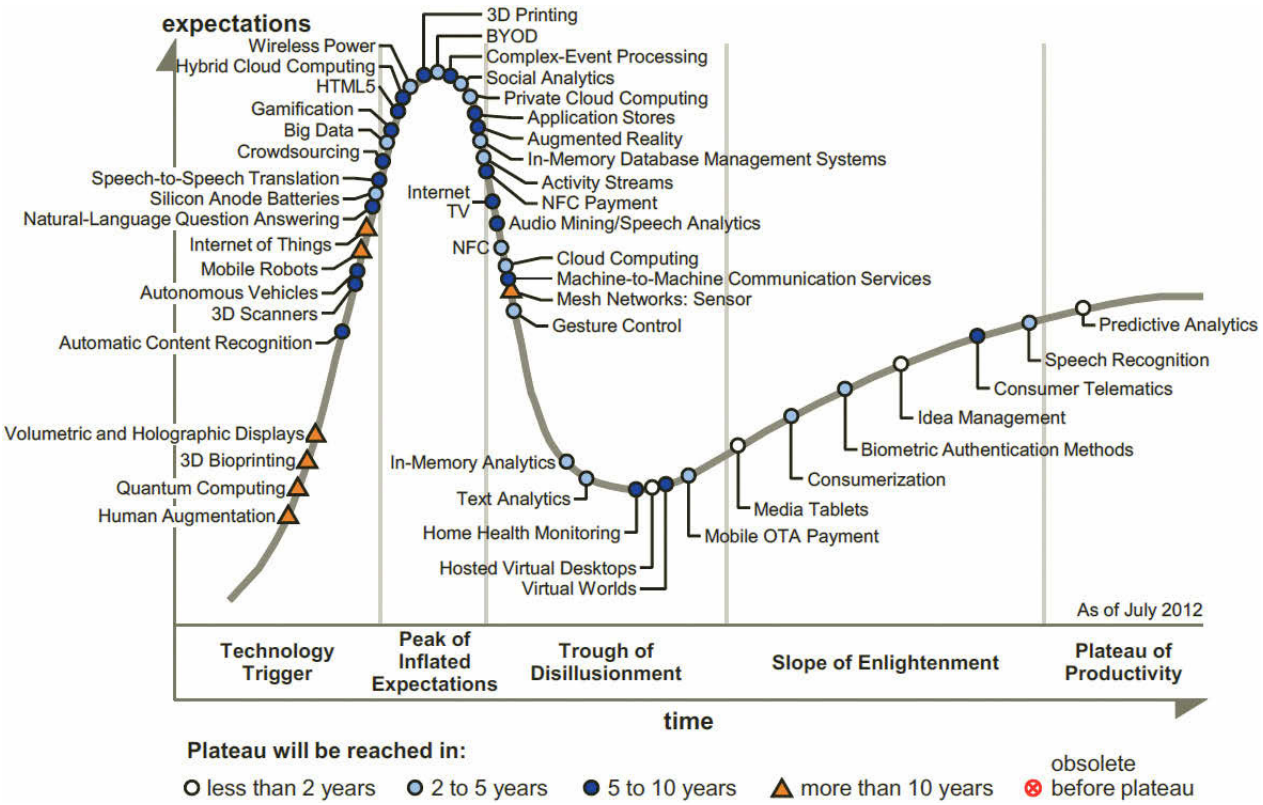
The most far-reaching development is the sky computing paradigm. Although the cloud metaphor originated from the widely used graphical icon for the internet symbolising the entire global network outside the user's computer, soon the term became applied to any computing service provided on the internet. The term sky computing extends the notion of cloud computing. It was coined for a setup in which clouds are combined into complex services, environments with workflows realising functionalities that exceed the capabilities of the individual services.

Language technologies are prime candidates for sky computing setups since they are often a component of complex applications such as services supporting knowledge discovery, business intelligence or text production. Taking into account the large number of languages, language variants and subject domains, a sky computing setup can provide a much larger number of language and task-specific workflows through service composition than a traditional software product. Small and medium technology enterprises will be able much more easily to enter the market, stay on the market and improve their services without having to cast all demanded service combinations into their product family or into a range of bilateral OEM partnerships.

3.6 THE FUTURE ROLE OF LANGUAGE TECHNOLOGY

In the next years language technology will play a major and decisive role, as explained and demonstrated by the discussion of megatrends and selected trends above.

The IT research and advisory company Gartner publishes the "Gartner Hype Cycle" every year. These studies are meant to provide strategists and planners with an assessment of the maturity, business benefit and future direction of more than 1,900 technologies, grouped into 92 areas [45]. Among the ones most prominently featured by the report are big data, 3D printing, activity streams, Internet TV, Near Field Communication (NFC) payment, cloud computing and media tablets. The Gartner analysts also mention several significant scenarios, that appear to be extremely promising on multiple levels but for which more enabling technologies are needed before they can be put into practice. Among them are "smarter things" and, most notably, "the human way to interact with technology". In fact, if we take a closer look at the 2012 hype cycle, reproduced in Figure 6, we notice that a total of 13 of the 48 technologies listed are language technologies, many of which are in the early "technology trigger" phase. Among the top emerging and key enabling technologies of 2012 and the coming years are, to list only a few, Automatic Content Recognition, Natural-Language Question Answering, Speech-to-Speech Translation, Complex Event-Processing, Social Analytics, Text Analytics and Speech Recognition. This assessment clearly shows that now is the time to invest in strategic research in the area of language technology and to go for a major, continent-wide push. One thing is certain: these technologies *will* come – they will be responsible for the biggest revolution in IT since the introduction of the graphical user interface and they will generate many jobs and countless business as well as social opportunities. Europe can now decide if it wants to play only a minor role, following the US and Asia, or it wants to move ahead and take the lead itself.



6: Gartner's 2012 Hype Cycle for Emerging Technologies [45]