



THE FORUM FOR EUROPE'S LANGUAGE TECHNOLOGY INDUSTRY

Findings on Language Technology Needs in Vertical Industries

Contents

LT for Security.....	3
Current challenges	3
Key findings.....	3
Opportunities	4
LT for Tourism.....	6
The Challenge	6
Key findings.....	6
Opportunities	7
LT for Healthcare	8
Key Issues.....	8
Key findings.....	8
Opportunities	9
LT for Manufacturing.....	11
Current challenges	11
Key findings.....	11
Opportunities	13

During 2013, LT-Innovate carried out a series of focus-group meetings with a number of vertical industries in Europe. The aim was to gauge awareness of the role of language technology in offering specific solutions for industry segments, and to learn how LT could anticipate and respond to emerging industry needs. This feeds into LT-Innovate's innovation agenda of helping LT buyers clarify their needs so that LT vendors can develop more appropriate and innovative solutions to language-related issues. The industries covered by this focus-group methodology are **Security, Healthcare, Tourism and Manufacturing**. We shall report below on each vertical in turn, breaking down the LT applications field into three domains for ease of exposition:

- **conversational interaction** covers speech technology applications;
- **digital semantics** covers the various creation and analysis applications involved in processing the meaning of content
- **multilingual access** identifies global language needs (cross- and multi-) for all other LT type operations.

On the basis of these and other focus-group conversations, LT-Innovate has inferred that one of the best ways to provide LT innovation to key European markets is to build flexible **collaboration platforms** focus on verticals. These are designed to enable a number of different technology vendors to join forces and supply sophisticated, adaptive solutions to the evolving requirements of each sector/subsector, usually by starting to develop an experimental solution with one major customer. This approach is based on the fact that, in many cases, vertical players know what they wish to achieve in the area of smart solutions for their end customers, but are not deeply enough informed about how LT (as a "new" IT sector) can help them do it. A collaboration platform, then, is a new kind of business services platform (not an *IT* platform in the usual sense of the term). It enables relevant LT players to work together to understand a company's needs, identify the necessary research and development required to upgrade existing technology, and supervise the flow of knowledge between customer, researcher and vendors (who may include non-LT components) to build an innovative solution via a collaborative industrial project.

A Language Technology-as-a-Business Service platform should therefore be able to deliver disruptive innovation that will:

- simplify the marketing of solutions by a given vendor,
- reduce the cost of LT solutions
- ensure interoperability of resources and applications
- scale solutions
- deliver upgrades and closely integrate them with other applications.

LT for Security

Current challenges

Security is a horizontal market that ranges over multiple business sectors, especially finance, government, energy, construction, defence, and legal. The management of risk-sensitive digital data and the risks involved in maintaining and protecting corporate assets are only the tip of the iceberg of the security management market today. The range of security issues is rapidly moving beyond logical or computer security into the areas of social media, Big Data, global regulations, and mobile access to information.

- Risk management and asset security in any sector are largely determined reactively by the power of the “enemy”. But businesses in general realise that proactive and predictive methods and strategies of risk management must be implemented in the short term to ensure higher risk control. Beyond standard business risk situations, there are numerous new phenomena emerging in the digital economy: for example, consumers are using social media to organise anti-business strikes; information privacy concerns are growing more acute (due to the easy circulation of private, confidential documents of all sorts); identity theft is rampant; and non-compliance with risk standards is raising insurance fees, and raising interest on debt re-financing.
- Security is therefore a multi-faceted target for LT solutions. However, the fact that speech interfaces are increasingly being used (e.g. in mobile phones) to negotiate access to secure sites, and that large quantities of textual information in multiple languages are required at all points in security workflows to solve problems and convey regulatory standards suggests that LT could play a role in simplifying certain critical tasks in the security galaxy.

Key findings

The role of conversational interaction

In the security market, the key technology application for voice/speech today is voice biometrics and the use of voice prints to identify and authenticate a user accessing a secure service. This is particularly critical in the mobile space where banks are gradually opening up their services to mobile users, and where office workers of all kinds are using their mobile phones to dictate messages and consult documents online. The field of voice biometrics, in other words, is already largely understood by key security players in the finance, energy, and legal industries, who are centrally concerned with business security issues.

Discussions did however reveal new opportunities for security departments across all the industries in question, which agreed that using speech input for document authoring would provide net gains in time and efficiency for document content creation and preparation. Speech recognition has been pioneered in the legal sector as a way for busy lawyers to accelerate the input of briefs and streamline the document production chain.

The role of digital semantics

Most of those working in the security sector appear to be aware of the power of text mining and related semantic technologies to discover information in Big Data repositories, identify sentiments on social media and hence manage brand images, and learn about the way information about security issues can circulate within different networks.

However, there appears to be much less awareness of the availability of smart authoring tools for those involved in producing documentation, and of the utility of semantic search tools to check and

prevent people who are trying to access sensitive data online or in corporate repositories, especially in the nuclear industry.

Semantic technologies (using concepts as search entities rather than just words) are being used, for example, to screen employees on whether they are entitled to access static databases, as well as tracking historical transactions, email records and the like to check on whether there are security breaches in a given department, database, or email stream. In the legal branch of the security marketplace, it is equally important to understand the role that semantic tools can play in identifying the existence or not of compliance with regulations, of the infringement of laws, and discovering relevant information from past case law. This latter is already a major focus of the legal profession. Yet all of these processes can be critical to establishing the legal status of many other actions in industry and business in general.

There was a felt need for developing an “attack graph” that would identify the entire range of different paths that might be used by anyone attempting to penetrate a given asset. The fact is that the scope of security has expanded with the arrival of social media and product traceability online, which both put pressure on security concerns. Big Data security will also begin to pose problems for companies. Document anonymisation (usually performed manually today) is another area where language automation can play a role, as sensitive documents increasingly need to be shared in a legal context without compromising named entities in the texts.

The role of multilingual access

Evidence gathered from meetings with security officers from various sectors (Legal, Finance, and Energy) suggests that most of them are not familiar with translation as a requirement in the security domain in general, nor with automated translation solutions. Security concerns for example involve tracking social media and related networks that carry private comments on brands, companies or services, and this requires multilingual access to social media content.

In the Energy sector, language requirements for monitoring threats change as new groups around the world begin to pose threats to industrial sites for political or other purposes. The multilingual analysis of emails to identify spam or security-centric issues is a further problem. At the same time, security departments are concerned that free access to translation automating services online raises problems of document confidentiality

Opportunities

Although some form of LT is often embedded in many corporate desktop applications used in various parts of security-minded businesses, the IT departments involved are often not aware of the functionality involved. There is widespread awareness of voice authentication and identity solutions across the industry.

On the other hand, there is very little awareness of the role of translation in improving information about and solutions to security touch points in companies. As companies themselves grow more international and multilingual de facto, there is a need to examine the specific kinds of solutions that multilingual access to information could pose.

- In the financial services market, LT can help ensure compliance with industry regulations, provide reliable voice authentication for mobile or e-banking.
- Legal services: LT can be used to accelerate the production of template-based real cases where data has been masked
- In telecoms, the provision of remote authentication by voice identification and even emotional voice monitoring to identify stress levels in users.

- In healthcare: identity verification, real time transactions to reduce communication failure and patient misunderstandings in treatment.

In general, LR vendors need to develop a compelling message around security applications and use this to learn more about the specific needs in the security dimensions of major industries. Often this will involve examining how LT can be effectively combined with other relevant security technologies in value-add packages.

LT for Tourism

The Challenge

Tourism is a major source of revenue and jobs in Europe, given the EU's role as a global tourism destination, albeit uneven in value across different countries. There is currently a sharp increase in Asian tourists and in special-interest and special age group tourism.

- Most importantly for LT-Innovate, tourism practices will be directly affected by the emergence of a new generation of digital natives, and the near-ubiquitous use of mobile devices (phones and tablets) as tools for accessing tourism information and conducting retail transactions in real time while on the move.
- The tourism market also intersects critically with the transport market, especially in the area of personal vehicles. Relations between consumer tourism information and in-vehicle information systems will become increasingly close, and even competitive.
- Tourism is obviously consumer-centric: this means that natural conversational transactions are core to tourist practices at every step in the value chain, from booking transport and hotels to seeking local information, listening to commentaries, and shopping.
- Conversational in this sense covers both voice and text content: the key feature is that the linguistic content involved is usually vernacular and informal. In addition, all tourist interactions leave a powerful data trail that can be leveraged in various ways by operators using multilingual tourism data analytics tools.

Language technology is therefore set to play a key enabling role in optimising certain aspects of the provision and management of the tourism experience, ensuring rich data understanding for the market, and simplifying cross-language communications between all parties in the psychologically high-stress situation of being a leisure-oriented stranger in a largely unknown cultural environment, where finding one's way is critical to well-being.

Key findings

The role of conversational interaction

The rise of the conversational interface on mobile and tablet devices will largely influence the future of tourist experience management. It is likely that much self-service information search will be carried out asynchronously via text/chat online rather than by spoken calls in a foreign country, but contact centre-type responses to tourist calls should be equally ready to handle spoken requests in any tourist language via the smartphone interface.

Speech output (TTS) to headphones or as alert messages in public spaces will probably become mandatory in a variety of emergency and other situations, ranging from in-vehicle driving directions and museum guide commentaries to tourist site information services accessible via tablets or phones, and services in railway stations and airports.

Spoken language recognition will be an added advantage in this context, so that speech translators that can handle either purchasing and information transactions or emergencies in the languages of the tourist/host will ensure that in-country services become a source of trust and comfort for tourists.

The role of digital semantics

Information is the life blood of the tourist market, both in the form of the input provided to tourists and in the knowledge learned about tourists from data concerning their behaviour on and offline (using various sorts of data analytics of comments pages and social media generally). Any functionality that

makes it easier and faster to create, edit, upgrade, recycle, quality-control, publish and translate information for and from tourists will be relevant to assuring better quality services, lower costs and greater competitiveness for tourist operators.

Any means for monitoring social media and other online opinion streams to discover highly specific trends and sentiments about specific tourist touch points will be equally important for planning and strategizing among tourism retailers, especially by analysing multilingual opinion streams. Accelerating and lowering the cost of translating original content will also be of paramount interest in this context.

The role of multilingual access

As in any multilingual context, tourism will benefit at every touch point in the value chain from robust translation support that can adapt intelligently to different transactional situations. These naturally include full customer language responses for website information (either pre-translated or performed automatically on the fly). The language range will need to be truly global in this vertical: access to vernacular content in all visitor languages, and provision of information to the same – major and minor languages alike. There can be no absolute reliance on the supposition that everyone speaks a lingua franca.

However, the strategic question in this multilingual access dimension is whether translation should be deployed as

- a) a device-specific seamless end service (i.e. the tourist is in control and uses apps/services that automatically provide a steady stream of the tourist's own-language information by on-the fly translation as necessary)? Or
- b) a delivered service developed upstream by the tourist facility – e.g. the localised web page – and delivered to the tourist's device as a packaged service via a standard workflow?

The translation industry in Europe is most familiar with the more linear solution (b), using service suppliers whose translation output is passed on to site clients. There may be a need to rethink the relationship between BYO devices and their multilingual capabilities to ensure that the translation function draws less on non-European free machine translation services, and more on European services that can deliver the appropriate quality and added value.

However, it may be that tourists prefer to use their familiar smartphone-specific translation service. This is an empirical question that the industry and LT-Innovate need to explore when developing innovative solutions for this industry.

Opportunities

LT-Innovate believes that tourism could be one of the priority verticals for a sustained LT innovation agenda. The size of the market, the fact that it ranges over every country and language in the EU, and that it is highly conversation-focused makes it a prime target for natural language understanding automation.

There are opportunities in all three application areas of technology, conversational, semantic and multilingual but especially in combining these dimensions of LT in innovative new solutions for ensuring information access for all parties.

A successful business services platform for this vertical could act as a showcase for the benefits of collaborative LT, and enable the sharing of best practices and technologies for other, similar enterprises.

LT for Healthcare

Key Issues

The on-going digitisation of the healthcare market in Europe (eHealth, digital health, telemedicine, etc.) is a major challenge for the health sector as a whole, but is already well under way. It will mean a radical change in work habits and knowledge management for doctors, clinicians, nurses, and many other hospital and para-medical jobs. Digitisation will also massively change the way in which patients and citizens understand and receive care, relate to the healthcare industry, and self-manage their health.

- For the LT sector, this new healthcare and medical regime will require an understanding of the impact of such technology drivers as mobile devices, cloud computing and interoperability on patient-driven healthcare (personalisation & empowerment, the mining and analysis of big medical data), and how social media can provide a new knowledge source for epidemiology, among other things – all of which are information (and therefore language) rich processes which will require fairly complex processing requirements.
- The shift from a “pre-digital” to a “digital native” population of patients and carers will vary across countries, but will ultimately lead to the emergence of sharable electronic health records for all, which in turn will call on big data techniques (including textual analytics) to provide insights for implementing smarter healthcare policies, and hence greater personalisation of care for individual patients. The digitisation of healthcare knowledge will also require a stringent privacy and data protection policy.

LT will therefore be called to play an enabling role in optimising solutions at specific touch points in the multiple tasks involved in data inputting, device manipulation, clinical conversations, searching health-records and knowledge sources, aiding in diagnosis and drug discovery, and in the longer term sharing of multiple forms of health information across all EU languages to improve and lower the cost of health policies, and help predict future health challenges.

Key findings

The role of conversational interaction

Technologically mediated spoken intercourse will come to play a key role as an interface medium for medical knowledge production because it is considered more ‘natural’ and closer to the patient/doctor mind-set. This includes clinicians dictating patient stories into their health record repositories, radiologists inputting analytic findings, nurses dictating reports, etc.

In a longer-term perspective, speech technology solutions will also play a risk-reduction role in video-meetings between medical staff (by recording and then transcribing the proceedings for reference and search), and also in tele-health situations where surgeons operate and control instruments and relate to other medical staff remotely.

There is already plenty of feedback about problems and failures in the marketplace that will eventually help build better solutions for the clinical story-telling case in the immediate future (e.g. the widespread recognition that clinicians dislike form filling). But as mobile devices become ubiquitous in a healthcare setting, enabling the recording and search of conversations to provide for eventual risk analysis, it seems fairly likely that the technological mastery of conversational speech as an interface and spoken content as a knowledge resource will help optimise specific workflow pressure points in the healthcare industry.

The role of digital semantics

LT-Innovate has identified two major areas where semantics is playing a vital role in enhancing healthcare knowledge management. One is in the pharmaceutical industry, where semantics-driven automated search can lead to the discovery of new ideas for drug research based on academic and other documents that would have taken too long to find by human examination of document and knowledge repositories alone.

Another is the use of LT components to aid medical specialists making diagnoses. In this case, fine-tuning diagnoses can be aided by automated searches over databases of symptoms to enable wider, deeper and more specific hypotheses to be examined in a given diagnosis.

There are almost certainly other areas in the domain of healthcare where semantically-enhanced search, classification of documents, and inference can be applied to provide more massive coverage of available medical knowledge, shorter lead times, and more accurate discovery of patterns in data relevant to public health monitoring or data collection for individual disease profiles. Semantics will also need to become more multilingual in scope and power.

The role of multilingual access

So far automated translation as an enabling technology has had a very limited impact in the healthcare industry. Hospitals and related institutions that make up the bulk of health organisations usually call on translation when they are dealing with patients who cannot speak any of the current languages (typically native + English) of the institution. This makes patient-doctor translation a non-systemic need, and is usually handled by human interpreters or in some other nonce way such as a shared second language.

But with the more fluid movement of populations (in tourism and migrant flows), translation may become a more systemic issue in clinical examinations, which in many cases could benefit from spoken translation support using high-quality automated solutions. From the medical side, there is a routine fear that valuable semantic nuance will be lost in machine translation, raising a barrier against optimum care and therapy.

It is clear that healthcare websites, social media dedicated to specific disease communities, and many other information and sharing resources would benefit from European-wide translation support. But the evidence suggests that there is no major *market* for translating large amounts of medical findings into all European languages as yet. The *professional* market for “medical” translation currently focuses on *medical devices and products*, a subset of the software or manufacturing localisation market, and the associated documentation (instructions for use (IFUs), informed consent forms (ICFs), training and user manuals, Patient-facing materials, Research material, Technical papers, Regulatory documents, Product literature).

Opportunities

The European Commission’s [Cloud-for-Europe initiative](#) will eventually provide a transformative environment for eHealth solutions in Europe. LT Innovate believes this cloud should be interoperable with any future Language Technology-as-a-Service solutions to enable automated local language coverage, content semantics, and speech cloud resources for all health management IT requirements across the EU. This interoperability between the eventual European Language Cloud and the Cloud-for-Europe will be critical for a maximally accessible, smart and communicative European health service to emerge.

Independently of a cloud solution, it is not clear that multilingual medical/healthcare platforms could easily be built as collaborative LT enterprises in the immediate future. Healthcare covers a highly fragmented archipelago of semi-interdependent domains, with different EU countries choosing

different regulatory and technical options. LT solutions and services therefore have to adapt to local conditions and choices made by local health authorities, even though interoperability is the ultimate goal for global health records and global health data. Large collaborative LT ventures might not be suited to this situation in the immediate future.

There is however a specific opportunity for inventing smart digital applications that can handle healthcare issues for people working or travelling outside their base countries – tourists and other migrants. These could make use of translation and other medically-informed language support, and would need to be trained on high quality resources carefully managed and sustained by a cloud-like language-resource infrastructure.

A further opportunity will come from the rapid rise in data capture from the IoT (internet of things) devices, social network interactions, and other big-data sources of personal health information. The task of analysing, applying discovery engines and addressing responses to such analyses – and eventually acting predictively on the basis of rich input - could largely be automated by integrating multilingual NLP and semantics into the analytic and visualisation solution. It would then be vital for the evidence and findings to be published (and translated where necessary) in simple clear language and channelled to medical services that need to act on such data.

LT for Manufacturing

Current challenges

Manufacturing is a major driver of European economies and an equally major employer. The main challenge for language technology that LT-Innovate has identified is to transform the growing mass of diversifying yet still siloed content (largely unstructured data) in multiple languages into an energy resource that can help speed up processes, render them more agile and more responsive to market demand. This covers all functions right along the chain from research and design through production and logistics to after-sales services.

- Production processes are becoming increasingly international, in the same way as distribution networks. This means that both customers and the workforce in a given company may be intensely “multilingual”. Although English is often used as a common corporate language to overcome this barrier, there are very large variations in individual competence. This can lead to considerable time and effort expended on repairing and enhancing internal communications.
- The same goes for access to critical information sourced in different languages. Current translation workflows tend to be expensive as more languages need to be brought into the mix as manufacturers globalise. And different functions need to have an adapted search and reliable ‘translate’ facility in order to access content in nearly real time. This demands a massive investment in multilingual terminology to ensure that crucial technical content is both up to date and properly understood by all.
- Due to the highly distributed nature of manufacturing plants for a given company, meetings are increasingly held online as videoconferences for all business functions from the design office to financial briefings. This cuts down on cost and carbon footprints but raises the problem of simultaneous translation for conference attendees unable to speak the common language sufficiently well. Providing integrated cross-language facilities in meetings would be a first step towards more effective information exchange and social cohesion within the enterprise.

Key findings

The role of conversational interaction

Conversational interfaces that enable spoken communication between remote colleagues or to control machines are obviously set to play a significant role at many points in manufacturing processes. Speech recognition is currently used in repair workshops and in warehousing where hands are already busy and voice communication and control can be effective. At the same time, almost any function in manufacturing companies that involves accessing information quickly via a computer or similar device could theoretically benefit from powerful speech recognition solutions.

More specific to widely distributed manufacturing organisations is the question of enabling project teams from different countries and languages to communicate efficiently during online meetings, increasingly via videoconference technology. Manufacturers have expressed the wish to see speech-to-speech translation, for example, developed as an enabler for online meetings. This is a major challenge as the solution would have to address not just the terminological details of the technical manufacturing domain but be capable of being used with any device from fully-fledged videoconference technology with telepresence down to a tablet or even a mobile phone during a quick project review meeting with team members.

The key demand here is that there should be a flexible set of basic language solutions that can adapt to the many different kinds of situations that occur across the manufacturing field. And that new solutions do not disrupt a sometimes ageing workforce by demanding new or unfamiliar skills.

An ancillary role for speech technology is also to be found in the increasing trend of designing conversational interfaces into the products that are made. This is especially true in the automotive sector but also in any almost any type of machinery (e.g. aircraft or in due course robots) that are typically controlled by humans.

The role of digital semantics

Digital content in manufacturing is growing exponentially, as it is in every vertical. This brings two key problems for end users in different manufacturing functions – quickly finding the right information from a broad range of document types, and ensuring that access is maximally cross-lingual.

Semantic search, enhanced by a toolbox of document classification, ontology, and taxonomy type solutions, will help the right people to access the right document at the right time, which is today an almost generic act in almost every field of work, and this is doubly true of manufacturing, due to the massive impact of compliance with rules and regulations on the design and production of almost every conceivable manufactured item. People working in such companies also need to be able to examine and discuss such documents either orally or textually simultaneously from different locations. Any advance in processing the meaning of documents or deriving a summarised report of the meaning of a collection of documents or other content items will be useful in this information access context.

As many large manufacturers tend to use SAP enterprise resource software to manage their processes, it would be advantageous to ensure that document-centric language technology can integrate smoothly into this framework.

The role of multilingual access

Manufacturers are centrally concerned by cross-lingual document production and product localisation (i.e. the interfaces, help databases, FAQs and related content that need to be adapted to the local language and culture of the end user of the product in question). They are therefore interested in using automated translation solutions to reduce translation costs and enable near real-time access to information. Those who have tested machine translation solutions for their various document requirements are inevitably dissatisfied by current results, and there is much room for improvement in this area. One key criterion for this language technology focus is to assure full “language compliance” of the content – i.e. the high technical quality of the translation.

A number of very large equipment manufacturers have experimented with controlled language as a means of standardising and simplifying the text of regularly updated documents so that a machine translation system can “learn” more easily how to translate the document accurately. Information quality can be a critical factor in many cases for ensuring compliance with end user expectations and supplier specifications, and hence reducing the number of post-sales support incidents.

However, manufacturers who have built and implemented controlled language checkers find that it is a long, costly process. Perhaps the critical need here is for intelligent tools that can detect problematic ambiguities in the natural language of user documentation and similar documents and signal it in some way to the translation process so that linguistic issues can be anticipated ahead of time.

Opportunities

One manufacturing executive described his ideal solution to cross-lingual and multilingual problems in everyday operations: a huge central database of information that everyone can access via their own local language interface in real time. LT-Innovate believes that it is worth listening closely to needs expressed in this way as they tend to coincide with the “ELC” cloud view of the future of language technology development proposed by other industries and policy makers.

Under pressure from globalisation, shorter development cycles and agile processes, manufacturers appear to require a new generation of LT tools and solutions that provide a more integrated approach to language-related tasks. At the same time, many LT vendors note that different branches of large manufacturers pursue independent technology acquisition policies that do not always fit into a standardised, cross-functional matrix vision of a given company’s needs.

There may well be opportunities, therefore, for the LT community to propose more integrated solutions around the central manufacturing information repository concept to overcome the typical silo-isation of resources, and enable all departments and branches to do more collectively with a single platform of language solutions.