Machine translation at Volkswagen: a case study

Jörg Porsiel

"The limits of my language mean the limits of my [working] world." (adapted from Ludwig Wittgenstein: *Tractatus Logico-Philosophicus*)

The Volkswagen Language Portal, available in the Volkswagen AG intranet since the summer of 2002, is a central reference point for all group employees for a broad scope of information on language in general – including the new German orthography with links to appropriate websites – as well as more specialized information concerning translation and interpreting. Furthermore, it provides access to a comprehensive multilingual, company-specific terminology database that currently has approximately 15,000 entries and also to a machine translation (MT) tool in German, English, French, Russian and Spanish. The portal and its contents are available in all of these languages.

Initial situation and framework conditions

The Volkswagen Group, with nearly 330,000 employees of whom slightly more than 50% are in Germany, sells its vehicles in more than 150 countries worldwide. Vehicles of the brands Volkswagen, Volkswagen Commercial Vehicles, Audi, Seat, Škoda, Bentley, Lamborghini and Bugatti are produced at 48 different production locations. In 2007, approximately 6.2 million vehicles were sold with sales of 108.9 billion euros. The VW foreign language management department in Wolfsburg manages approximately 30,000 translation orders in up to 40 languages and approximately 1,200 interpreting orders every year.

The initial impulse for the provision of the language portal, the terminology database and MT came from several things – for example, from the increasing necessity, in the age of internationalization and globalization, to increase the quality and speed of text processing and translation processes through a standardized corporate language, as well as the continued work on comprehensive terminology management in the field of technical documentation. These objectives are to be achieved through the installation of a simple tool for processing e-mails, the number of which has increased enormously in the last few years both within and outside the group. Furthermore, the huge increase in the text volume of technical documentation due to the extension of the vehicle model range and the addition of new markets has to be managed.

Several billion valid, non-spam e-mails are sent every day worldwide. On average, an employee of one of the companies listed in Fortune 500 – and Volkswagen is one of those companies – receives 80 e-mails a day. According to estimates, 70% of the employees in a large company in Germany can lose one hour every working day processing incomprehensible e-mails. This includes correspondence in a foreign language. The increasing number of e-mails does not necessarily mean an improvement in levels of communication or in the exchange of information.

Inefficient exchange of information can also result from foreign language e-mails being misunderstood, not understood

Jörg Porsiel has worked in the field of translation, terminology management and foreign language corporate communication since 1990. He has been at Volkswagen AG Foreign Language Management since 2002.

at all due to insufficient language knowledge or being answered late due to the time needed for human translation. MT can accelerate this process considerably. In split seconds, text quantities ranging from one line to several hundred pages can be translated. Time saving is the primary objective for the use of MT.

Alongside the continued improvement in language quality and in the speed of communication in the company, the content in the portal should promote a lasting awareness of language in general and also the particular importance of a standardized and clear corporate language.

Economic benefit

The language portal and especially MT are currently being used by employees to reduce the time needed for daily routine tasks, such as the processing/translation of e-mails, minutes from meetings or reports. This is primarily a question of creating so-called "indicative translations" or raw translations. Although the quality of these raw translations can vary enormously, depending on the quality of the source texts, the result of the translation is enough for most users to be able to decide what to do next within seconds based on their own specialized knowledge.

The primary benefit for Volkswagen results from the constellation of information requirement, sourcing and processing in conjunction with the considerable time saving achieved with MT. In times when product cycles (time-to-market) are becoming ever shorter and globalization continues, the acceleration of information sourcing and processing and the refinement into knowledge are of great importance in securing advantage over the competition. The time factor is decisive particularly for companies working in all time zones. The application of MT for routine tasks such as the reading, answering or forwarding of e-mails leads to a previously unknown acceleration of foreign language communication and decision-making processes, linked at the same time with a considerable and lasting reduction in costs for processing and translation.

MT in use

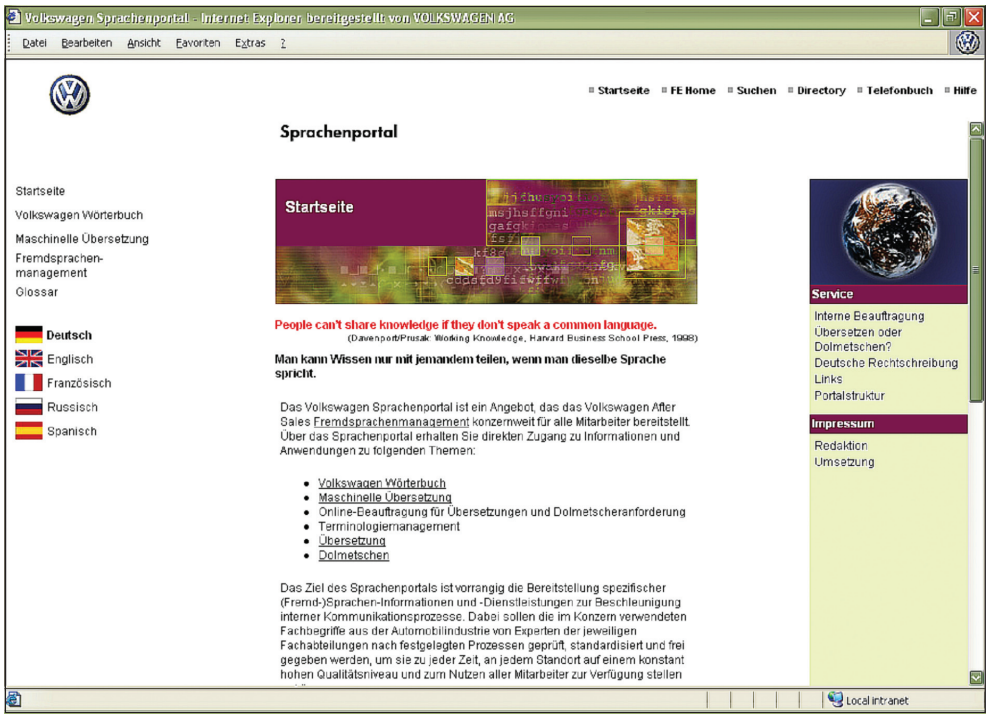
The MT tool in use in the language portal was selected after a series of comprehensive tests. One clear objective was – and is – not to replace human translators through MT or to use MT in a productive environment, for example, for the translation of texts that have been translated by human translators up to now. MT is used nearly exclusively for the translation of texts that the users have otherwise had to translate themselves, using bilingual dictionaries or other aids. This process is comparatively time consuming.

There is a rule-based system (RBMT) made by Lucy Software and Services GmbH for all of the language pairs on offer. Statistical machine translation (SMT) or hybrid systems, a mixture of SMT and RBMT, that have been coming on to the market over the last few years were either not yet ready for the market or not available at all mid-2002 in German, English and Spanish, when Volkswagen carried out the tests.

Key data

The language portal and with it the MT function are available for the whole group worldwide. After a targeted extension of the offered languages to include Russian (February 2008) and French (July 2008) as well as the integration of over 40,000 VW-specific terms (primarily from vehicle technology and engineering), an average of 7,000 MTs are currently being carried out every day. A translation can be simply one word or a text of more than 100 pages. Overall, this means a translation volume of approximately 1,500 standard pages per day, whereby – and this must be emphasized – the quality becomes of secondary importance. The primary and exclusive objective is speed and thus time saving.

The language portal content has been increased considerably over the last two years, not least through the growth of the terminology database which is now available in 20 languages with approximately 15,000 German terms and contains exclusively group-specific concepts. Furthermore, the scope of MT was extended to include French and Russian and the VW terminology database integrated in the five portal languages in order to improve translation results. The content



The Volkswagen Language Portal in five languages.

of the language portal has been translated completely into the five current portal languages.

What MT can achieve

If MT is to be used sensibly, one must first consider where its application would be advantageous – but also what MT cannot do and, primarily, why it cannot do it. The portal provider must also promote awareness of this fact.

The use of MT is worthwhile if the following prerequisites are given: there must be a specific corporate terminology in the largest possible scope and of the best possible quality in both the source and target languages. Each terminological concept must be clearly identified and also allocated to a particular field. Ideally there should be an interface between the terminology database and the MT tool, or there should at least be regular updates of new terminology so that the quality of the MT results can be continually improved. Alongside saving time, this will lead to an increase in acceptance from the users, who are in general initially skeptical about the use of MT.

The users must also be made aware of the overriding economic and personal benefits. Furthermore, the users must be encouraged to use this terminology in their daily work. It would also be very useful to implement machine authoring tools in fields such as technical documentation in order to increase efficiency, reduce translation costs and promote language awareness in a lasting fashion.

The quality of translation results in MT – much more so than in human translation – depends on the quality of the source texts. Users occasionally complain about the results of MT. Two items of evidence are used in this context: 1) The translation of individual terms does not function. "I entered *xyz*, but did not receive a translation." Alternative: ". . . but received a completely incorrect translation." 2) Some users are of the opinion that they have found the best method of testing translation quality: reverse translation. Due to a lack of language awareness, they believe that language is the same as mathematics in its structure and function. In the same way the cross check is used in mathematics, the text that has just been translated is translated back into the original language. If the result is not completely or at least 90% identical with the original text, the user will automatically deem it to be a bad translation. Both points are proof of the deficit in user awareness or rather of how much work there is still to be done concerning understanding of language in general and of MT in particular.

One of the main sources of mistakes in MT – alongside insufficient terminology, software problems and other technical problems – is the users themselves. Some examples: due to a lack of feel for language and also usually no knowledge of what MT can and cannot do, users regularly try to translate individual terms – as with a dictionary or a terminology database. This will work sometimes, but normally it does not. Users will also often enter incomplete sentences, some with incomplete or even incorrect punctuation (primarily at the end of a sentence), incorrect (one might even say arbitrary) spelling or with unknown or self-created abbreviations. There is also a high frequency of texts that are largely made up of a language mix of German and English, for example, which makes it impossible for the computer to process them.

What MT cannot achieve

The majority of users are unaware – at least at the beginning – of what MT can do and what it cannot do, the latter being particularly difficult to communicate. One element of the learning process must therefore be to make clear to the users what MT cannot do, namely, read between the lines, guess, interpret, assume, associate, paraphrase non-translatable text and also be tolerant of mistakes.

Source text with mistakes:

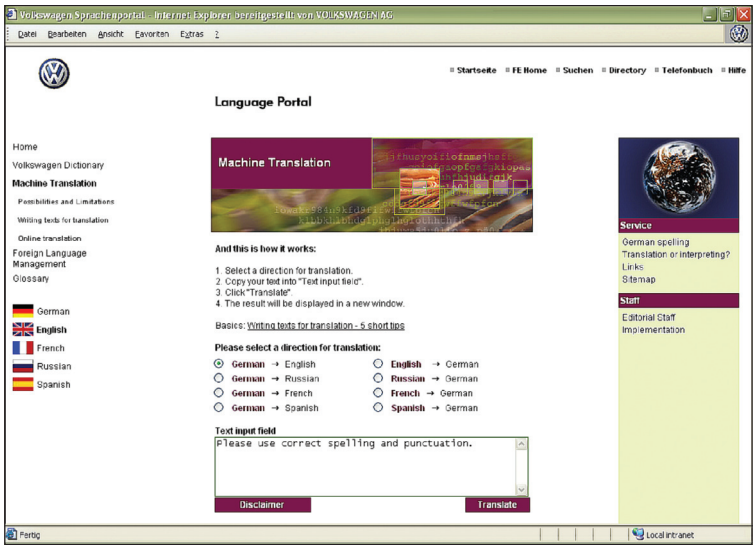
habe u.g. HWMotor 4J0.367.832 hinsichtl. ENBaulage an drei Bsp. aktueller Neuprojekte gegengeprüft. demnach ist dieser HWMot. für uns zur Anfrage

Result:

[u.g] has. [HWMotor 4J0.367.832] with regard to [ENBaulage] at three [Bsp]. current new projects' tested. accordingly this [HWMot] is. for us for the inquiry

Suitable source text using VW terminology:

Der neue Passat CC steht unter anderem mit einem starken und agilen V6-Motor mit 220 kW (300 PS) zur Verfügung, der serienmäßig mit Doppelkupplungsgetriebe DSG und Allradantrieb 4MOTION ausgestattet ist (Kraftstoffverbrauch: 10,1 l/100 km; CO2-Emission: 242 g/km). Im neuen Coupé kommt dazu ein innovatives System zum Einsatz: die Adaptive Fahrwerksregelung DCC.



Machine translation in the Volkswagen Language Portal.

Result:

The new Passat [CC] is available among other things with a strong and agile V6 engine with 220 kilowatts (300 hp), [which | who] is equipped TO direct shift gearbox and to four-wheel drive 4MOTION in series with direct shift gearbox (Fuel consumption: 10,1 l/100 km; CO2-[emission | exhaust]: 242 g/km). In the new Coupé an innovative system is used for this purpose: the Adaptive Running Gear [Regulation | Arrangement] DCC.

MT in the internet

Unlike the use of MT in a firewall-protected corporate intranet, the application of MT tools in the Worldwide Web brings with it risks that the vast majority of users have, up until now, largely underestimated or have not even considered: industrial espionage. Now, a large number of websites offer MT services "from any language to any other language" – a tempting offer at first as it is nearly always free of charge. But the potential user should always ask himself or herself a few questions beforehand. Where is the server? Who has access to it? What happens with my data? Who is the provider? Why are they offering this service? What are the interests of the provider? As these questions can hardly be answered by the user, one should consider one's options carefully before using such services – above all, if the text content is confidential.

User reaction

The Volkswagen Language Portal is currently registering approximately one million hits per month, and the trend is rising strongly. Although there are, as already mentioned, some users who have a negative view of MT itself and of the translation results in particular, the vast majority of feedback is extremely positive. The positive responses clearly indicate the assistance that MT can provide for daily tasks. A large number of users even offer assistance for the improvement of translation results by providing subject-specific terminology and even want to import it themselves into the appropriate dictionaries.

A look ahead

The most important aim for the future and above all extended application of MT is the growth in the scope on offer and the areas in which it can be used. Alongside the continued growth of the corporate terminology database – also from non-vehicle technology specific fields (such as economics, law and insurance) – this includes the implementation of a direct interface between the terminology database and MT as well as the addition of further languages such as Chinese, Portuguese, Czech or Polish.

The extended application of MT could be implemented in fields with a highly regulated or standardized texts such as those in diagnosis devices or similar systems. The absolute prerequisite in this case, however, would have to be that the source text has been checked by a machine authoring tool, as otherwise the post-processing work would be too lengthy, which would mean that costs and advantages would not add up.

Finally, the objective must be to continue to focus on the importance of "language" in any group working at a global level and to emphasize the benefits for the individual, the flow of communication and the exchange of information. The results would be not only a noticeable improvement of the internal communication among employees of different locations throughout the world, but primarily also a considerable increase in the quality of communications with suppliers and customers. **M**

A German version of this article will also appear December 15 in MDÜ *Magazine, the quarterly magazine of the German translator's industry association BDÜ.*