Machine Translation: The Linguistic Epiphany

by Laura Rossi

Manager Language Technology Solutions,
LexisNexis

LexisNexis®

Machine Translation: The Linguistic Epiphany

"There is no need to do more than mention the obvious fact that a multiplicity of language impedes cultural interchange between the peoples of the earth, and is a serious deterrent to international understanding."

This is how Warren Weaver, American scientist and mathematician (July 17, 1894 – November 24, 1978), opened his Translation Memorandum in 1949. He continued on, in great modesty, to say, "The present memorandum, assuming the validity and importance of this fact, contains some comments and suggestions bearing on the possibility of contributing at least something to the solution of the world-wide translation problem through the use of electronic computers of great capacity, flexibility, and speed. The suggestions of this memorandum will surely be incomplete and naïve, and may well be patently silly to an expert in the field — for the author is certainly not such." [1].

Needless to say that his thoughts on the possibility of translating natural languages through computers to solve worldwide communication among different languages were not at all 'silly.' Crossing cycles of enthusiastic belief and disillusion, machine translation has been and still is the subject of a great deal of patents, proof of the fact that the problem of cross-language communication is far from being solved, and that technologies promising to do so can indeed help the progress of humanity and, not secondarily, be of significant economic value.

Are you also wondering what the latest advances and applications have been in the machine translation space in China, Japan, Korea, Germany and France? Would you also like to know where machine translation is moving to? I for one, am pretty curious—namely because this has been my area of professional expertise for quite a long time, and because of a life-long passion for languages and automation. In order to do so, I am going to use machine translation itself in LexisNexis® TotalPatent®.

LexisNexis TotalPatent strives to offer the most complete collection of full text patent authorities and adds to it English machine translations, enabling its customers to run cross-authority searches by using a single language.

Take-away #1: The business choice of translating the entire full text into English has to do with the desire of both maximizing accuracy of translation (consequently of search results), and of offering customers ready-to-read translations of the entire documents at their retrieval.

[1] (1949) Warren Weaver: Translation. 17 July 1949. Available at: http://www.mt-archive.info/Weaver-1949.pdf. Site last accessed on 6 October 2015.

To spend a moment on matters of accuracy, machine translation systems produce more accurate results when the words are translated in context, because only context can help the machine disambiguate polysemous words – that is to say words that bear more than one meaning. Such context can be derived by either co-occurring words (as in our case, since we translate the entire text) or predefined taxonomies, such as domain classifications or domain dictionaries. Whereas statistical machine translation systems rely mostly on the first technique and learn the correct translation by 'seeing' many examples of usage in context, rule-based systems normally use domain dictionaries for disambiguation, which are compiled by terminologists. In rule-based systems, the machine will need to be told up front which dictionaries to use dependent on the domain field of the document to be translated. This explanation, however, is a little simplistic. Nowadays, in fact, statistical machine translation training more and more often uses domain clues to help disambiguation in addition to the textual context, in particular where data is sparse and unequally distributed across domains.

Let's have a look at a couple of examples. The Italian translations of the word 'mouse' in zoology or informatics are respectively 'topo' and 'mouse.' For achieving high accuracy in machine translation, we would need to have either the word in context ('A transgenic mouse having a phenotype characterized by the substantial absence of mature T-cells' versus 'Optical mouse having an integrated camera') or an indication of some kind that the single keyword 'mouse' we are trying to translate into Italian refers to either zoology or informatics.

The example below, which was run in Google Translate, exemplifies this matter once again. The English word 'tablet' has multiple meanings, which are possibly expressed by different words in other languages. Take French, for example:

1.	[for writing – stone, wax etc]	tablette f
	[- pad]	bloc-notes m
2.	[pill]	comprimé m, cachet m
3.	[of chocolate]	tablette f
	[of soap]	savonnette f
4.	[plaque]	plaque f (commémorative)
5.	COMPUTING	tablette f[2]

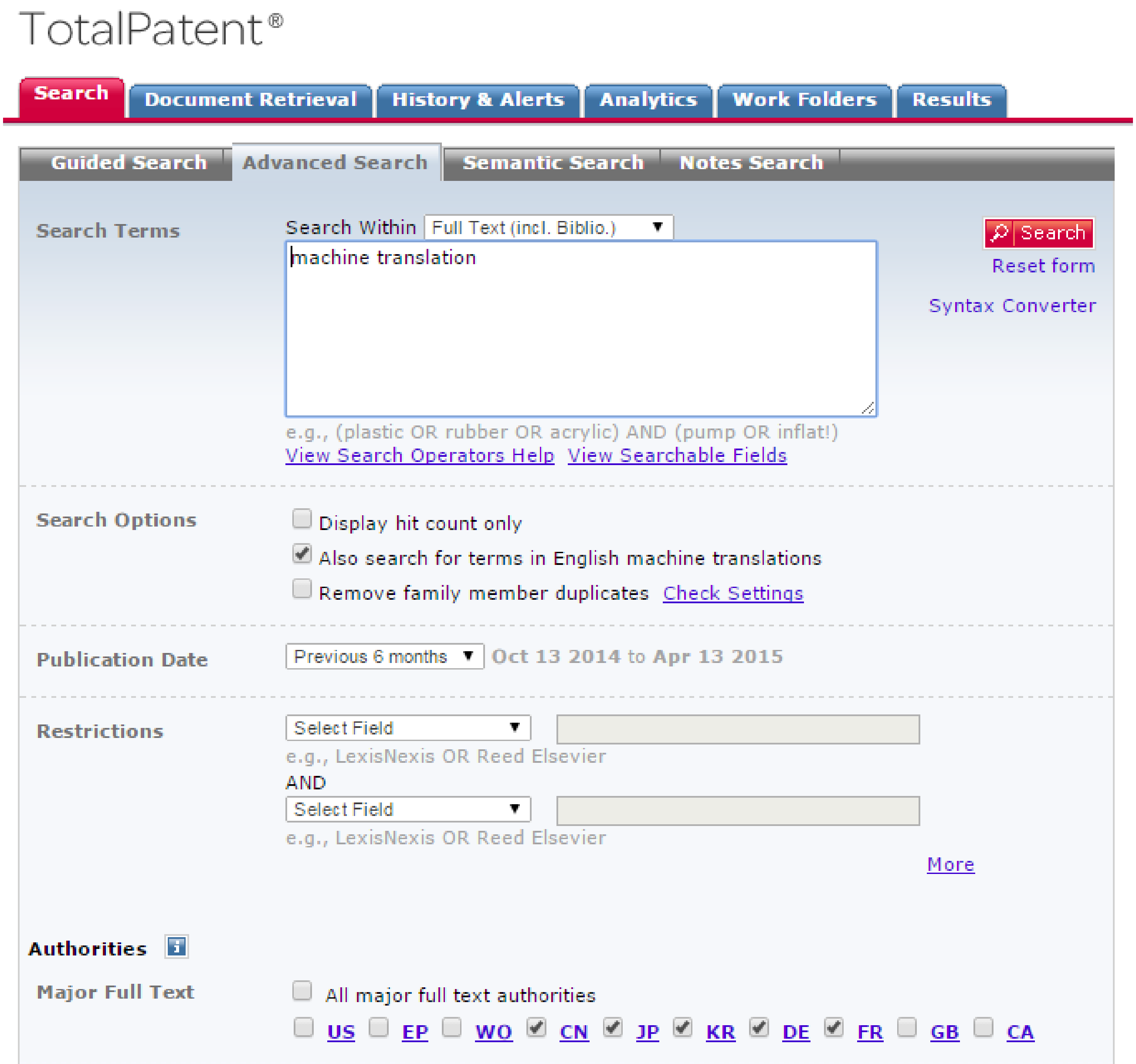
[2] See: http://www.larousse.fr/dictionnaires/anglais-francais/tablet/616857. Site last accessed on 6 October 2015.

As you can see in the table below, the more context that is added to the keyword, the more reliable its translation becomes.[3] Short strings with no further descriptive context will have one or the other translation dependent on how frequently they appear in their singular form, plural form, or are accompanied by articles or prepositions in the training data set (unless the translation system is specifically designed to cover only a particular domain, such as medicine, informatics, food industry etc.).

English	French Machine Translation via Google Translate
tablet	tablette
tablets	comprimés
a tablet	une tablette
the tablet	la tablette
the tablets	les comprimés
with the tablet	avec la tablette
the tablet with	le comprimé avec
a tablet for writing	une tablette pour l'écriture
a tablet of chocolate	une tablette de chocolat
a commemorating tablet	une plaque commémorative
a keybord for my tablet	un clavier pour ma tablette
a medicinal tablet	un comprimé médicinal
Vitamin B compound tablets containing...	Comprimés de vitamine B composés contenant ...
Portable tablet with keyboard and wireless mouse...	Tablette portable avec clavier et souris sans fil ...

[3] Samples translated via Google Translate on 8 June 2015.

Back to the search. Having TotalPatent at my disposal, I ran the following simple query in English against the authorities of my interest. The query searches both in original English and machine-translated text.



Search Terms — Search Within: Full Text (incl. Biblio.) ▼

machine translation

e.g., (plastic OR rubber OR acrylic) AND (pump OR inflat!)

Search Options
- □ Display hit count only
- ☑ Also search for terms in English machine translations
- □ Remove family member duplicates

Publication Date: Previous 6 months ▼ Oct 13 2014 to Apr 13 2015

Restrictions: Select Field ▼ e.g., LexisNexis OR Reed Elsevier AND Select Field ▼ e.g., LexisNexis OR Reed Elsevier

Authorities

Major Full Text: □ All major full text authorities

□ US □ EP □ WO ☑ CN ☑ JP ☑ KR ☑ DE ☑ FR □ GB □ CA

While I am expecting most of my results to belong to IPC Class G06 (COMPUTING; CALCULATING; COUNTING), to my surprise the 232* retrieved results spread across all sections and, in particular, a variety of classes in IPC Sections A, B, E, G and H. A little puzzled, I start looking at some of the results belonging to different IPC Sections, to finally have a linguistic epiphany. Even if, in my perception of the world, the word 'translation' mostly refers to **(1)** '*the process of translating words or text from one language to another,*'[4] the same word has important alternative meanings in the technical and scientific space, among which **(2)** '*Formal or technical: the process of moving something from one*

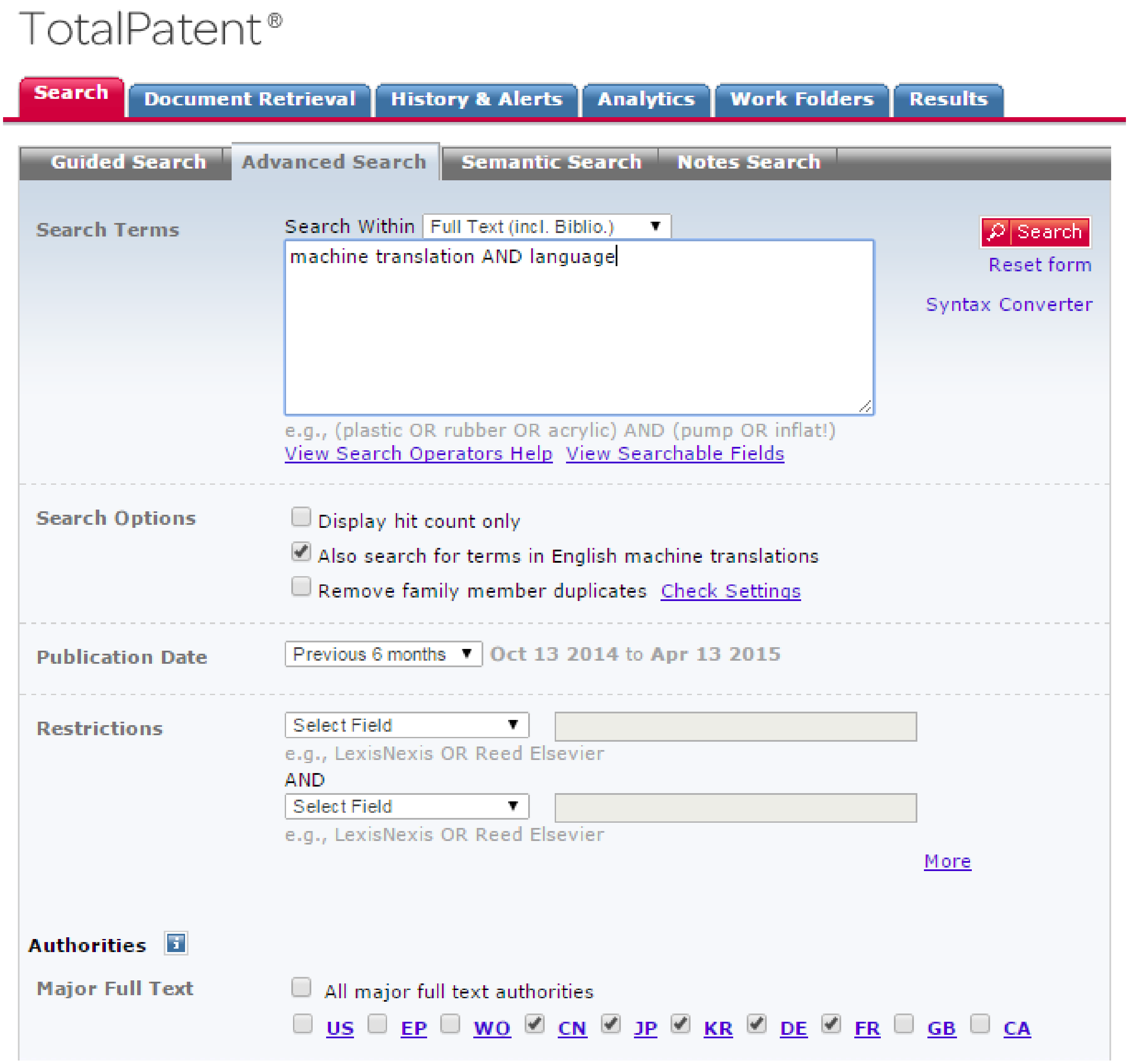
* Please beware that running the same query now will give different results, because of continuous updates and corrections happening in the database.

[4] This and the following definitions of the word 'translation' originate from: http://www.oxforddictionaries.com/us/definition/american_english/translation. Site last accessed on 6 October 2015.

place to another,' **(3)** *'Biology: the process by which a sequence of nucleotide triplets in a messenger RNA molecule gives rise to a specific sequence of amino acids during synthesis of a polypeptide or protein,'* and again **(4)** *'Mathematics: movement of a body from one point of space to another such that every point of the body moves in the same direction and over the same distance, without any rotation, reflection, or change in size.'*

In particular, in its technical sense of 'process of moving something from one place to another,' it is not difficult to imagine that this process could be automated and driven by a 'machine,' in which case the term 'machine translation' could pop up in other IPC Sections – in particular IPC Section B (PERFORMING OPERATIONS; TRANSPORTING) with a very different connotation.

I can narrow down my search by disambiguating the word 'translation' with the addition of the extra search string 'language.'



Search Terms

Search Within: Full Text (incl. Biblio.)

machine translation AND language

e.g., (plastic OR rubber OR acrylic) AND (pump OR inflat!)

Search Options

☐ Display hit count only
☑ Also search for terms in English machine translations
☐ Remove family member duplicates

Publication Date: Previous 6 months Oct 13 2014 to Apr 13 2015

Restrictions

Select Field
e.g., LexisNexis OR Reed Elsevier
AND
Select Field
e.g., LexisNexis OR Reed Elsevier

Authorities

Major Full Text

☐ All major full text authorities

☐ US ☐ EP ☐ WO ☑ CN ☑ JP ☑ KR ☑ DE ☑ FR ☐ GB ☐ CA

I am now presented with 155 results. While the majority of them belong to the IPC Section G, and in particular Classes 05 (CONTROLLING; REGULATING), 06 (COMPUTING; CALCULATING; COUNTING), 07 (CHECKING-DEVICE) and 10 (MUSICAL INSTRUMENTS; ACOUSTICS), there are still a couple of results not matching my expectations. One is a Chinese patent belonging to the A61K48 Group (Medicinal preparations containing genetic material which is inserted into cells of the living body to treat genetic diseases; Gene therapy): CN104411338A. When examining the document I indeed find out it mentions 'machine translation' in the following paragraph:

> 本文部分地提供编码目标多肽的多核苷酸、初级构建体和/或mmRNA，所述多核苷酸、初级构建体和/或mmRNA已经被设计成改进以下中的一项或多项：在组织中的稳定性和/或清除率、受体摄取和/或动力学、组合物的细胞通路、与翻译机器的接合、mRNA半衰期、翻译效率、免疫逃避、蛋白质产生能力、分泌效率(适用时)、循环的可及性、蛋白质半衰期和/或细胞状态、功能和/或活性的调节。
>
> *This text to provide part of the polynucleotides encoding the target polypeptide, the primary construct and/or mmRNA, polynucleotide a plurality of, mmRNA primary construct and/or have been designed to improve the following in one or more of: the stability of the in the tissue and/or clearance rate, receptor uptake and/or dynamics, composition of the cell passage, the joint of the machine translation, mRNA half-life, translation efficiency, immune evasion, protein production ability, secretion efficiency (when applicable), the accessibility of the cycle, protein half-life and/or cell state, function and/or modulation of the activity of.*

Although I cannot read Chinese, I can see the paragraph mentions mRNA (Messenger RNA), which is also present in definition **(3)**, and I assume that 'translation' is used here with that connotation. Whether this translation really happens by 'machine' or whether the entire translation is completely correct, only a human translator expert in the domain can tell. **Machine translation** is known to possibly make mistakes, because it **is not a perfect science** and, as in our case, it learns by statistics.

Take-away #2: Even when machine translation sounds perfect and all words have been translated, it is advisable and wise to remain 'skeptical' and never assume its full correctness. Machine translation is in no way a substitution for a human translation. Despite all customization efforts, this technology uses probabilistic models that are, by their nature, imperfect and can result in mistranslations, omissions or additions of non-pertinent information in the target text. Even human translators make mistakes – even if, most likely and trustingly, in a much lower measure.

Coming back to my search results, the chosen disambiguation strategy did not work too well in this case, because the word 'language' appears in a paragraph highlighting some body symptoms, including 'language' related issues.

> 症状包括但不限于发育停滞、腹泻、呕吐、黄疸、流鼻血趋势增加、头小畸形、震颤、共济失调、自残行为、精细运动协调障碍、语言缺陷、痉挛和/或卷心菜样气味。
>
> *Symptoms include, but are not limited to, development of stagnation, diarrhea, vomiting, jaundice, nosebleeds increasing trend, head small malformations, tremor, ataxia, behaviors, fine movement obstacles coordination, language defect, spasms and/or odor cabbage type.*

Having discarded this document from my search result list, I am intrigued by the five hits belonging to Section H, and specifically:

Publication	IPC Main | First	Assignee
CN102651750B	H04L29/08	ALIBABA GROUP HOLDING LTD
CN102422639B	H04N7/15	CISCO TECH INC
CN104170293A	H04J3/16	GOOGLE INC
KR1020150003879A	H01L21/265	VARIAN SEMICONDUCTOR EQUIPMENT ASSOCIATES, INC.
CN104333072A	H02J7/00	安徽省新方尊铸造科技有限公司

The translation for one of the assignees is only available in Chinese (at the time of search). Some people may wonder why we do not use machine translation for proper nouns of inventors or assignees. The reason is that the same character sequences used to build proper nouns can also be used for common nouns, and, as a consequence, no sensible translation can come out of a machine translation system (unless it was trained on names and their translations). In addition, many times the names in Chinese, Japanese or Korean are already a translation or transliteration of Western names. Applying a second layer of translation can lead to hilarious (and undesired) results (see how Mr. 弗朗茨·约瑟夫·欧池, possible translation for Mr. Franz Josef Och, PhD inventor listed in patent CN102150156B, Distinguished Research Scientist at Google between 2004 and 2014 and one of the main authorities in the field of machine translation, becomes 'Franz Josef European Pool' in this unsuccessful attempt of 'back translation')[5]. At LexisNexis, machine translation is only used for the translation of 'textual and technical' elements: titles, abstracts, descriptions and claims.

Take-away #3: Machine translation is not the right tool for translation of proper nouns. For translation of assignees, we rely on other sophisticated strategies, such as statistical lookup techniques.

A closer analysis of these five hits shows the following: In document CN102651750B, 'machine translation' is mentioned multiple times as a tool that helps the realization modes of the invention, which is a 'Method, system and device for providing Web page information.' The invention is not about machine translation itself, but can make use of machine translation. In CN102422639B 'machine translation' has a more prominent role and appears both in the description and claims of this 'System and method for translating communications between participants in a conferencing environment.' Google's application CN104170293A mentions 'machine translation' in claims 11 and 20 as an integral part of the invention. Korean document KR1020150003879A talks about 'machine translation' and 'language' (in 'programming language,' however), but it looks like the terms are used in a different context and the hit is not relevant to my search. Finally, CN104333072A is about a practical application for speech/voice recognition.

[5] Sample translated via Google Translate on June 16, 2015

‘Machine translation’ is mentioned in the section discussing Background Art: ‘Speech recognition technology with other natural language processing techniques such as machine translation and voice synthesis technology, can construct a more complex application, such as a voice to the voice translation.’

The interesting thing is that **none of these five documents reports the term ‘machine translation’ in the text of either title or abstract**, not even the ones in which ‘machine translation’ is mentioned in the claims!

Take-away #4: By translating the complete full text (all descriptions and claims) of its patent collections into English, LexisNexis has created the possibility for its customers to run cross-authority queries by using a single natural language, and to find relevant hits even when the textual information that is searched is exclusively available in the description or claims.

Having explored these five results belonging to IPC Section H, I can say with certainty that at least the first three give me some information about possible applications of ‘machine translation.’ They all have the commonality of belonging to Class H04 (ELECTRIC COMMUNICATION TECHNIQUE), and of showing, through their later classifications, a link to G06 as well (COMPUTING; CALCULATING; COUNTING), the Class in which I would have expected the majority of the results of my query to be classified. Chinese application CN104170293A, which had been assigned Class H04 by the Chinese Patent Office, has even been classified differently by the European Patent Office by assignment of CPC Class G06. The remaining two results, the ones I personally also found less relevant, do not have any reference to Class G06, even in their later classifications.

Take-away #5: Full-text search in machine-translated text does not substitute classification-based search, but is a powerful complementary tool to validate searches performed through other criteria and a way of ‘expanding’ possible results beyond the boundaries of our expectations.

Coming to the results in IPC Section G, the three hits classified under G05B19/418 (CN104423341A, CN104423347A and CN104423350A), explore ‘machine translation’ in the context of speech translation in a voice recognition system (hence the advanced classification under G10L) for remote controlling of household electrical appliances.

The two hits under G07C9/00 (CN104331968A and CN104331967A), instead, are quite intriguing. They both refer to a system for allowing vehicles access to a private parking space, based on respective recognition of the owner's voice or fingerprint. But what does this have to do with 'machine translation'? Whereas I can find a correlation between 'machine translation' and 'speech recognition,' I still wonder what this has to do with 'fingerprint recognition.' A closer look at the paragraphs in the two documents reveals something interesting:

	CN104331968A	CN104331967A
1	语音识别技术，也被称为自动语音识别 Automatic Speech Recognition，(ASR)，其目标是将人类的语音中的词汇内容转换为计算机可读的输入，例如按键、二进制编码或者字符序列。	指纹识别技术，也被称为自动指纹识别Automatic Speech Recognition，(ASR)，其目标是将人类的指纹中的词汇内容转换为计算机可读的输入，例如按键、二进制编码或者字符序列。
2	与说话人识别及说话人确认不同，后者尝试识别或确认发出语音的说话人而非其中所包含的词汇内容。	与说话人识别及说话人确认不同，后者尝试识别或确认发出指纹的说话人而非其中所包含的词汇内容。
3	语音识别技术的应用包括语音拨号、语音导航、室内设备控制、语音文档检索、简单的听写数据录入等。	指纹识别技术的应用包括指纹拨号、指纹导航、室内设备控制、指纹文档检索、简单的听写数据录入等。
4	语音识别技术与其他自然语言处理技术如机器翻译及语音合成技术相结合，可以构建出更加复杂的应用，例如语音到语音的翻译。	指纹识别技术与其他自然语言处理技术如机器翻译及指纹合成技术相结合，可以构建出更加复杂的应用，例如指纹到指纹的翻译。
5	早在计算机发明之前，自动语音识别的设想就已经被提上了议事日程，早期的声码器可被视作语音识别及合成的雏形。	早在计算机发明之前，自动指纹识别的设想就已经被提上了议事日程，早期的声码器可被视作指纹识别及合成的雏形。
6	而1920年代生产的”Radio Rex”玩具狗可能是最早的语音识别器，当这只狗的名字被呼唤的时候，它能够从底座上弹出来。	而1920年代生产的”Radio Rex”玩具狗可能是最早的指纹识别器，当这只狗的名字被呼唤的时候，它能够从底座上弹出来。
7	最早的基于电子计算机的语音识别系统是由AT&T贝尔实验室开发的Audrey语音识别系统，它能够识别10个英文数字。	最早的基于电子计算机的指纹识别系统是由AT&T贝尔实验室开发的Audrey指纹识别系统，它能够识别10个英文数字。
8	其识别方法是跟踪语音中的共振峰。该系统得到了98%的正确率	其识别方法是跟踪指纹中的共振峰。该系统得到了98%的正确率。
9	到1950年代末，伦敦学院(College of London)的Denes已经将语法概率加入语音识别中。	到1950年代末，伦敦学院(College of London)的Denes已经将语法概率加入指纹识别中。
10	1960年代，人工神经网络被引入了语音识别。	1960年代，人工神经网络被引入了指纹识别。
11	这一时代的两大突破是线性预测编码Linear Predictive Coding (LPC)， 及动态时间规整 Dynamic Time Warp技术。	这一时代的两大突破是线性预测编码Linear Predictive Coding (LPC)， 及动态时间规整Dynamic Time Warp技术
12	语音识别技术的最重大突破是隐马尔科夫模型Hidden Markov Model的应用。	指纹识别技术的最重大突破是隐马尔科夫模型Hidden Markov Model的应用。
13	从Baum提出相关数学推理，经过Labiner等人的研究，卡内基梅隆大学的李开复最终实现了第一个基于隐马尔科夫模型的大词汇量语音识别系统Sphinx。	从Baum提出相关数学推理，经过Labiner等人的研究，卡内基梅隆大学的李开复最终实现了第一个基于隐马尔科夫模型的大词汇量指纹识别系统Sphinx。
14	此后严格来说语音识别技术并没有脱离HMM框架。	此后严格来说指纹识别技术并没有脱离HMM框架。
15	尽管多年来研究人员一直尝试将“听写机”推广，语音识别技术在目前还无法支持无限领域，无限说话人的听写机应用。	尽管多年来研究人员一直尝试将“听写机”推广，指纹识别技术在目前还无法支持无限领域，无限说话人的听写机应用。

The two paragraphs only differ by the usage of 语音 ('voice') versus 指纹 ('fingerprint'). The rest of the text is identical and clearly refers to the Background Art of speech recognition technologies and not of fingerprint recognition – as if the patent drafter simply performed a 'search and replace' action in the patent regarding the application of fingerprint recognition technology to this usage. 'Machine translation' is mentioned in sentence #4, which respectively translates as follows:

> *CN104331968A*: Speech recognition technology with other natural language processing techniques such as **machine translation** and voice synthesis technology, can construct a more complex application, such as a voice to the voice translation.
>
> *CN104331967A*: Fingerprint identification technology with other natural language processing techniques such as **machine translation** and fingerprint synthesis technology, can construct a more complex application, such as a fingerprint to the fingerprint of the translation.

The second translation really does not sound good. 'Fingerprint recognition' is not a 'natural language processing technique' and cannot be combined with 'machine translation' to build the described application. Clearly the source text does not help us create a good translation here.

Being very well aware of the issue, LexisNexis has invested years in improving text digitization through optical character recognition technologies and creating intermediate layers of preprocessing between the original source text and the machine translation, so as to minimize the chances of error. Low quality of the source text due to misspelling, faulty content and terminological inconsistencies, however, cannot be easily detected nor corrected before the text is submitted to the translation workflow, meaning that these errors could be propagated from source to target.

There are still some results left for our analysis: Eight of the hits belong to IPC Group G10L15 (Speech recognition). Not difficult to imagine that 'machine translation' might be used as a technology in combination with speech recognition as seen for previous results, or, alternatively, be mentioned in the Background Art as one of the advancements in natural language processing that is most close to speech recognition.

Take-away #6: The primary variable influencing the quality of a machine-translated text is the quality of its source text. According to the famous GIGO principle ('garbage in, garbage out'), computers, since they operate by logical processes will unquestioningly process unintended, even nonsensical, input data ('garbage in') and produce undesired, often nonsensical, output ('garbage out')[6].

[6] See http://en.wikipedia.org/wiki/Garbage_in,_garbage_out

I am also curious about the remaining ones, the ones in Class G06. I will try to summarize some of the innovations in and around machine translation by listing some of the titles of the granted patents in the set of retrieved results:

Country	Publication	Titles	Title is:
JP	JP5694893B2	Device for selecting the optimum translation, translation selected model learning device, method, and program	**MT**
JP	JP5666937B2	Mechanical translation device, mechanical translation method and program	**MT**
JP	JP5652824B2	Text input device, a translation device including the same, and a computer program text input method	**MT**
JP	JP5650440B2	Method and device for weight learning program identity, N-best scoring device, N-best rerun king device, their	**MT**
JP	JP5646529B2	Translation candidate presentation device, and translation candidate presentation program translation candidate presentation method	**MT**
JP	JP5632213B2	Mechanical translating device and program	**MT**
JP	JP5615476B2	The phrase presenting program translation, and the translation phrase presenting method translation phrase presenting device	**MT**
CN	CN102272754B	Customized language model	**MT**
CN	CN102708147B	A technology method for identifying new word of terms	**MT***
CN	CN102637167B	Multilingual translation method	**MT***
CN	CN102033879B	A Chinese names method and apparatus for identification of	**MT***
CN	CN102708098B	Coherence constraint based on dependence of the bilingual words automatic alignment method	**MT***
CN	CN102945232B	Training-corpus quality evaluation and selection method orienting to statistical-machine translation	**HT**
CN	CN102799579B	Statistical machine translation method with error self-diagnosis and self-correction functions	**HT**
CN	CN102662936B	Chinese-English unknown words translating method blending Web excavation, multi-feature and supervised learning	**HT**
CN	CN102165435B	Automatic context sensitive language generation, correction and enhancement using an internet corpus	**HT**
CN	CN102609410B	Authority file auxiliary writing system and authority file generating method	**HT**
CN	CN101814067B	System and methods for quantitative assessment of information in natural language contents	**HT**
CN	CN102799578B	Translation rule extraction method and translation method based on dependency grammar tree	**HT**
CN	CN102609409B	Online translation method, device, system and server	**HT**
CN	CN102591859B	Method and relevant device for reusing industrial standard formatted files	**HT**
KR	KR101475795B1	Media object query submission and response	**HT**
KR	KR101453937B1	CJK name detection	**HT**

MT = Machine Translation | **HT** = Human Translation
MT* = Machine Translation in the process of being replaced by Human Translation

Language model improvements; methods for quality assessments; name detection; new word detection; evaluation and selection of training corpora; methods for extractions of translation rules; applications in cross language retrieval; combined application with speech recognition. Lots of work done in or for the Asian Market. Among the assignees some big names of the data and computing business: Microsoft, Google, IBM, Yahoo, Amazon, Baidu, Xerox, Fujitsu, Toshiba and plenty of universities, revealing how the technology is still pretty much anchored to the academic world and big research centers. I think the question of my, to say it in Weaver's words, 'patently silly' search query has been answered.

A couple of things are still worth mentioning.

It is thanks to machine translation that I could find much of this information and that I could read and understand, on a global level, what is going on, especially so soon after the official date of publication. As soon as text in the original language is received by LexisNexis, a machine translation of all the text elements (title, abstract, description, claims) is generated. Once an official human translation for titles and abstracts is delivered by the patent offices, the human translation replaces the machine translated text. The translation marked with 'MT'* in the table above are in the process of being replaced with human translations at the moment of consultation of the content repository.

Take-away #7: Machine translation represents a powerful instrument to unlock information in a foreign language and to do it quickly. In patent search, timeliness of information is key. No matter how much we believe in the value of this technology, however, we recognize its boundaries and we do give priority to human translations where available. In the end, humans still do a better job at this.

It is also thanks to machine translation that we can explore the entire patent landscape at once in our quest for Prior Art. If querying 'machine translation' AND 'language' in only English language authorities (i.e., US, EP, WO, GB, CA, AU, IN, IE) with no date restriction delivers 9,474 hits (as of date of search), expanding the search to CN, JP, KR, DE and FR makes the number rise to 13,176. In our current globalized world, ignoring these extra 3,702 results does not seem like an option.

Take-away #8: The choice is whether to confine ourselves to the 'familiar' and the 'comprehensible' or to welcome the use of machine translation, with its natural imperfections, to explore possibly relevant knowledge published in languages that are unknown to us. While doing this, we keep in mind the 'motive' and 'practical objective' behind the use of this technology, which Prof. W. John Hutchins reminds us of:
'The motive is the removal of language barriers which hinder scientific communication and international understanding. The practical objective is the development of economically viable systems to satisfy a growing demand for translations which cannot be met by traditional means'[7].
From side of development, instead, we focus on the challenge he talks to us about: 'to produce translations as good as those made by human translators.' As much as this remains a utopia at the present time, it is in the end what really drives the machine translation community's constant quest for improvement.

LexisNexis® TotalPatent®:

TotalPatent is the world's largest collection of searchable full-text and bibliographic patent databases. It allows users to access extensive patent databases, including the full text from 32 patent authorities, bibliographic and abstract data from 68 authorities for a total of 100 authorities. All of this content is searchable, both in the language of publication and in English language machine translations, as well as images, legal status, citations and patent family data and compressed and searchable PDFs. No other patent research solution in the world offers that level of comprehensive reach and international scope.

7 (1987) W. John Hutchins. Prospects in machine translation. MT Summit: Machine Translation Summit. Manuscripts & Program, September 17-19, 1987, Hakone Prince Hotel, Japan; pp. 48-52. Available at: http://www.mt-archive.info/70/MTS-1987-Hutchins.pdf. Site last accessed on 6 October 2015.