The patent challenge

Is Japanese patent translation really that difficult? Well, yes, says Trevor Wright, an expert in this highly specialised and technical field – but it's a fascinating linguistic challenge. Here he explains why

Trevor Wright has a BSc in pharmacology, lived and worked in Japan for 11 years, then studied for an MSc in Medical Biotechnology in the UK. Following graduation, he worked as a senior patent analyst and translator for Thomson-Reuters and is now a freelance medical and patent translator at The Kanji Foundry (www.thekanji foundry.com). He has also developed an e-learning website (www. thekanjifoundry learning.com) for people wishing to teach themselves Japanese to English patent and medical translation.

After teaching English in Japan for 11 years, I returned to the UK to study for an MSc. While studying, I saw an advertisement in the *New Scientist* for a Japanese-to-English patent translator at a London publishing company, and applied for the job. Why not? I had nothing to lose. I bluffed my way through the interview and they offered me the job. I'd never translated much before then and had certainly never seen a patent. How hard could it be?

I actually learned patent translation on the job while working as a patent analyst and translator for this publisher. I looked at the way patent documents had been translated by others, and learned from them.

Patents are descriptions of inventions and ways for inventors to protect cutting-edge technology, which, mostly, has never before been made public. Innovative companies are prolific inventors and rigorously protect everything they develop. Samsung, for example, has filed over a quarter of a million patents to date. At the heart of a patent lie the claims which define the invention and which must be worded (and translated) meticulously. If the patent document is well written, there is no ambiguity in the text, but ambiguity is often written into patents so as to avoid narrowing the scope of the invention.

. Some words and phrases are preferred over others by different patenting authorities. The translator must always be clear about the differences between 'consisting of' and 'comprising', the distinction between 'that' and 'which', the need to understand the difference in wording of dependent and independent claims, and the requirement for the use of 'a' over 'the' (rendered more difficult due to

'Ambiguity is often written into patents so as to avoid narrowing the scope of the invention'

the absence of definite and indefinite articles in Japanese).

Each claim – and there can be thousands of them in a single patent application (US2003173072, for example, has 8958 claims) – may only be a single sentence irrespective of length. Claim 1 in US5408417 has 1228 words, with 40 commas and 23 semi-colons. Patents can be also be very long; US6314440 contains 3,300 pages.

Online databases make research easy. When I first started patent translation, there were no online resources available as there are today, necessitating a trip to the British Library in London each week. Searching took hours.

Today, thankfully, all the patent documents a translator needs can be retrieved from the comfort of their desk via their favourite browser. The European Patent Office maintains a searchable database of over 80 million patents (both applications and granted) from over 90 countries, in many languages, and the Japanese Patent Office offers over 8 million online documents. They are freely downloadable and are copyright free. Searching takes minutes; preparing documents for translation only a few minutes more.

Is it really that difficult?

Is Japanese to English translation really any harder than any other language combination? Well, given that Japanese has three scripts with more than 50,000 characters between them, no plurals, no relative pronouns, no spaces between words, no capitals, no future tense and little in the way of punctuation, I'd have to say yes. I'll elaborate on just a few of these points here.

Kanji, hiragana and katakana: three scripts make it easier. Oddly, three scripts in addition to the frequent use of English words make Japanese a lot easier to read than might be imagined. There are around 50,000 kanji (the Japanese name for Chinese characters (Hanzi) or ideographs); a good translator and an educated, literate Japanese person will be familiar with around 4,000 of these. Most kanji have more than one reading or pronunciation. The two syllabaries, hiragana and katakana, each have 46 symbols, each syllabary containing the 46 sounds that can be used to show the pronunciations of kanji. This means that it's possible to write a word, for example the word for disease, 4 ways; in kanji 病, in hiragana (びょう), in katakana (ビョウ) and in romanised script (byō) with the diacritic (macron) giving the 'o' a long vowel sound. Katakana is useful to represent words imported from English and other foreign languages, as well as for exporting Japanese words like マンガ (manga) and アニメ (anime).

The lack of relative pronouns can

be tricky but 'that' and 'which' present problems in English too. Japanese lacks the two important relative pronouns 'that' and 'which', so there are no relative clauses in Japanese. This may be a good thing as the use of these two words often confuses native English speakers.

Henkan misu

The 'henkan misu' is a form of typo. Typos occur in all languages. They are usually easy to spot and most word processors have built-in spell checkers, which offer at least some help. When inputting Japanese text, the word is typed phonetically using a Japanese kana or QWERTY keyboard. The spacebar is then used to convert the input (in hiragana) into kanji. If each kanji character had a unique pronunciation, conversion would be easy, but they don't. Shō, for example, typed on an English keyboard as 's h o u', will conjure up over 90 kanji, either as single characters or compounds (two or more kanji), with the same sound. The spacebar is repeatedly pressed until the correct character is highlighted in a list on screen, then Enter is pressed to select it. If Enter is pressed too soon, the wrong kanji character is entered. This is the henkan (conversion) misu (mistake). It occurs from time to time in patents and can require a fair amount of lateral thinking to figure out.

There are no spaces between words and punctuation may be of little help. When a translation agency asks for a quotation, they often ask how much I charge per word. Well, I'd have to translate the entire document first before I would have any real idea of the number of words in it. The cost of Japanese translation is usually quoted as a price/1000 characters. I wish all translation agencies and clients would grasp this!

Grammatical quirks

There are no plurals but plurality is inferred from the context. Although there are no plurals as such, the context often indicates whether more than one thing is being referred to. Plurality can be indicated by collective nouns like 'pair', 'group' and 'set' and there are words meaning 'some', 'several' or 'a few'.

Counting things presents many problems as there are over 100 'counters'. In Japanese grammar, counting of objects is suffixed with a kanji character to indicate the nature (often the shape) of what is being counted. As we may say 25 'head' of cattle in English, so Japanese has similar notation. Flat items like sheets of paper are suffixed with 枚 (mai), books with 本 (hon) (which can also be pronounced pon or bon depending on the preceding word as some consonants are subject to euphonic change, notably the plosive ones) and animals with 匹 (hiki or biki or piki) for small animals and 頭 (tō) for large animals. Birds and rabbits are counted in the same way (羽, wa) but fish are counted differently (they

---

'Japanese has three scripts, no plurals, no relative pronouns, no spaces between words, no capitals, no future tense and little in the way of punctuation'

---

take 尾, bi, which, logically, means 'tail') regardless of their size. In my field, a strand of DNA takes the same counter as a bottle of beer, both being cylindrical in shape.

There is no future tense. The present and past tenses can be clearly expressed but the future tense is not so easy. Japanese has a subject-object-verb structure so the overall tense of the sentence isn't clear until the tense of the final verb.

Punctuation and romanisation

There are no capitals – it's all in the brackets. Words can be capitalised in English to emphasise their importance. Titles, headings and subheadings can all be formatted with capitals and/or bold and/or italic font to indicate a hierarchy of importance within the text. Japanese often uses kagikakko, a kind of half square bracket (「 and 」) which are actually single quotation marks used when citing reference material, for example. Double quotation marks look like 『』. Other brackets, that is, ( ), { }, [ ], < >, 〈 〉 and 《》 are also used. Patent documents also use lenticular 【 】 brackets to indicate the claims and section headings. The use of brackets in this way can

sometimes make things simple. There is no need for bold, italic, capitalised or underlined titles or headings in a larger font size. It's all in the brackets.

The interpunct, also known by the Unicode consortium as the katakana middle dot, is a punctuation mark which looks like a floating dot (·) and is the equivalent of a forward slash (/), using the same key on the QWERTY keyboard. It is used for interword separation and gets translated as both 'and' (most commonly) or 'or' although it can also be used as a comma or a decimal point. It is also often used as a bullet point at the head of a list.

Before macrons were available in font sets, the long vowel sounds for ō and ū were denoted in rōmaji (the depiction of Japanese text using the English alphabet) as the digraphs ou and uu. Before word processors, these sometimes appeared as circumflexed vowels, ô and û, which were readily available on QWERTY typewriters. In fact, three systems of romanisation exist. The Hebon-shiki (Hepburn system) devised in 1885 by James Hepburn is still the most common, although the official system advocated by the Ministry of Education since 1937 is the Kunrei Siki (spelled using the Kunrei system), based on the older Nihon Siki. Older textbooks and dictionaries, particularly those prepared by various ministries of the Japanese government use the Kunrei system. For example, a modern romanised dictionary will render the suffix 状 as jō (meaning -like), while older dictionaries spell this as zyô.

Gemination (or consonant elongation) occurs in Japanese too. The character っ (little tsu) has no sound at all, but it indicates a double consonant and is used to elongate the sound. It is essential for the translator to bear gemination in mind when looking up words in dictionaries and typing words on a keyboard. They are easily missed.

These then are just some of the problems faced by Japanese to English translators and I haven't even mentioned genitives, morphemes, graphemes, the multi-tiered wooden box and hot water bucket readings of kanji, phonosemantics, phonomimes and psychomimes (words expressing a state of mind), of which there are thousands!