Blitz Latin Publications:

The Problem with Plaga

The ambiguity of Latin and its consequences for machine translation.

William A. Whitaker (McLean VA, USA)
and
John F. White (Wokingham, England)

This paper was first produced in late 2004 for publication in a well-known classical magazine. The magazine subsequently declined to accept it for unstated reasons, and it is therefore made available to users of Blitz Latin as a useful overview of what to expect from this popular machine translator. The paper has been very slightly updated, and a new appendix added.

Update 12 May 2006.

The Problem with *Plaga*

The ambiguity of Latin and its consequences for machine translation.

by William A. Whitaker (McLean VA, USA) and John F. White (Wokingham, England)

ABSTRACT

We describe how the ambiguity of Latin words renders the machine translation of Latin texts into English much harder than that of the derived west-European languages. Some possible solutions are outlined, and have been tested on the commercial translator 'Blitz Latin'.

INTRODUCTION

Latin is an ancient and primitive Indo-European language that is, by consequence, a language whose words are 'over-loaded'; that is, many words have multiple, unrelated meanings. It does not appear to have developed much before its speakers were exposed to another ancient, but much richer, language, that of the Greeks. Latin adapted for its own use very many Greek words. Some overlapped with existing Latin words.

A good example is the Latin word *plaga*, which can be derived from three early Greek stems [1]. The first (from the Greek to weave or entwine) led to the word *plaga*: 'net' or 'trap', although a second early Latin meaning was also 'curtain'. Another Greek stem gave the meaning 'region/tract of ground/zone', while a third Greek stem (to strike) provided the fourth form of *plaga*: blow (or wound, plague, scar, stroke or related meaning).

The four forms of *plaga* could be partially differentiated by pronunciation, although these could not be distinguished in the written form: *plăga*: net; curtain; region. *plāga*: blow. Even the Romans appear to have tired of this ambiguity, for *plaga* (curtain) appears by classical times as *plagula*.

Latin is also an inflected language, where the ending of the word conveys information about how the word is to be used. But the inflections are not unique either, so that, for example, the Latin word *reginae* might mean 'of the queen', 'to the queen', 'the queens' or 'O queens'. If we apply the ambiguity of the inflections to that of the stems, we have now up to twelve permutations for *plagae*: 'of the net/region/blow', 'to the net/region/blow', 'the nets/regions/blows' and 'O nets/regions/blows'.

When the Romans imported numerous Greek words into their language, they left a surprising omission. The Greeks had an inflected definite article ('the') that qualified nouns and aided in separating the possible alternative inflections of the noun. The Latin language lacks both the definite and indefinite ('a') article, so that the phrase *rex amat reginam* could mean 'the king loves the queen' or 'a king loves a queen'. Classical Roman writers often qualified nouns with adjectives such as *quidam*: 'a certain king loves a certain queen', but this was done for reasons of style rather than for grammatical simplification.

The net effect of the ambiguity of meaning for Latin stems, and the re-usage of common inflections in different ways, is to make Latin one of the most ambiguous of all languages to translate. This was understood even by the Romans, whose speech ('Vulgar Latin') contained

many more prepositions than the written form [2]. The prepositions controlled certain types of inflection; for example, ***ex*** takes the ablative case while ***in*** can take the accusative or ablative case. Vulgar Latin also made much use of the adjectives ***ille***, ***illa***, ***illud*** (‘that one’) to qualify nouns. In later centuries ***ille*** and ***illa*** would form the basis of the words for ‘the’ [i.e. il, la (Italian), le, la (French), el, la (Spanish); the neuter forms were dropped], in the Romance languages. The better-educated Romans found it still less ambiguous to write in Greek. However, the elegance and compactness of pure Latin for expressing grand thoughts and concepts ensured its popularity for monuments, speeches, history and poetry.

All languages have ambiguity of meaning for some words. In English, we might say ‘He hit a ball with a bat’. A reader unfamiliar with sport might wonder whether the ball had been struck with a shaped wooden stick or with a small flying mammal. However, Latin is particularly well endowed with this translator’s problem, through its use of over-loaded stems and over-replicated inflections. It is especially remarkable how many Latin words can be construed as a noun or as a verb, again a legacy of the language’s primitive origins. For example, the word ***reges*** can be translated as ‘the kings’ or as ‘you will rule’. The common origin is obvious.

Another consequence of the primitive nature of Latin is that its ablative inflection is used to describe conditions that in modern English would be identified as ‘with’, ‘from’ and ‘by (means of)’. In English there is a substantial difference between the phrases:
‘he leapt the fence with his horse’;
‘he leapt the fence from his horse’;
‘he leapt the fence by means of his horse’.
The Romans, however, intended the ablative case to be used as an adverb: ‘he leapt the fence horse-ily’.

If the inflection on the end of a Latin word conveys important information about how the word is used, then not to know the meaning of the word (and thence its inflection) can have a catastrophic effect on the interpretation of a Latin sentence. The meaning of inflections on those words that are known may be adjusted to try to make sense of the sentence without considering the contribution of the unknown word. If the reader strikes out as ‘unknown’ any single word in the preceding sentence, the meaning of the remaining English words remains unaltered. By contrast, consider the short phrase ***regis urbs pulcra est***: ‘the city of the king is beautiful’. If the word ***est*** is mis-spelled (perhaps as ***esst***), the sentence becomes ‘O beautiful city you rule [esst].’ To emphasise further the ambiguity of Latin, ***est*** also means ‘eats’: ‘the beautiful city of the king eats’.

The translation of Latin text requires absolutely that the reader should have knowledge of every word in each sentence. Anything less may cause mis-assignment of inflections and incorrect word ordering. From this it follows that the most important requirement for a machine translator of Latin is a sufficient vocabulary. If the words of a text are missing from the dictionary, the correct translation can never be supplied. This rather obvious statement of fact conceals a major problem of automatic Latin translation: the correct missing Latin word may be replaced by an overlapping ambiguous alternative, so that a noun may replace a verb or vice-versa.

There have been several attempts to create machine translators for Latin with a highly restricted vocabulary. While successful on simple test sentences, they fail hopelessly when presented with real, ambiguous, Latin texts. Suppose that our short Latin dictionary has the words ***rex, regis*** (king) in its dictionary, but to avoid ambiguity we have not added the verb ***rego, regere*** (to rule). Then the short sentence ***reges urbis reginam*** (‘you will rule the queen of the city’) becomes ‘kings of the city queen’, a legal, but nonsensical, translation.

A further complication for the Latin translator is that the language does not differentiate upper and lower case, as can be seen in any inscription. Thus proper names, themselves often used to describe trades, just as in English the name Smith is derived from the blacksmith’s trade, cannot readily be separated from the trades. For example, we have the proper name ‘aper’, indistinguishable from ***aper*** (a boar). However, later copyists of ancient texts took it upon themselves to capitalise the first character of proper names, so that an edited text will commonly refer to the proper name of ‘Aper’.

LATIN WORD ORDER

Latin retains another surprise for the unsuspecting modern reader, accustomed to the subject-verb-object word order of sentences in most modern European languages. The writer of Latin places the words in order of their emphasis. Thus the subject, object and verb of a sentence can be, and usually are, placed anywhere within it. The consequence is that Latin could not be translated word-for-word into English even if the words themselves were unambiguous. The word order needs rearrangement too. Latin textbooks advise the translator to identify first the subject noun of a sentence, and then its matching verb. You do not find this instruction in books about the translation of modern French, German, Spanish or Italian into English!

Two examples illustrate the essential problem to a translator:
rex amat reginam. ‘the king (subject) loves the queen (object)’.
reginam amat rex. ‘the queen (object) loves the king (subject)’.
Both sentences have the identical meaning in Latin, but a word-for-word translation of the second sentence will be confusing to a casual reader.

In the early Christian period of the Roman empire, when persecution was rife, a frequent expression was the phrase ‘against the followers of Christ’. The variants in English, Italian and Latin should be contrasted:
Latin: ***contra Christi cultores***.
Italian: contro i discepoli di Cristo.
English: against the followers of Christ.
The reader will observe that the number of Latin words and their order differ from those of Italian and English, which are word-for-word translations of each other. Moreover, ***cultores*** can also mean ‘cultivator’ or ‘inhabitant’. The roots of modern Italian lie in Latin, but the language has expanded to reduce ambiguity (Italian is a ‘rich’ language), more prepositions have been added (‘di’) and inflections exist only to separate masculine and feminine words. The English language has dropped even this distinction.

MACHINE TRANSLATION

The difficulties of translating by machine from one language to another have been recognised almost from the earliest days of computing. It was soon discovered that human translators

have a large advantage over their machine counterparts: the humans could use their extensive general knowledge to help separate alternative, ambiguous, meanings. Indeed, so disheartening was this discovery for the early pioneers of machine translation that many soon abandoned the effort. Others, however, persevered, with the result that machine translation can be useful in three general areas:
1. Within narrow, often highly specialised, fields, within which the limitations can be made negligible. For example, to provide user instructions for a manufacturer's range of televisions, after human translators have corrected the defects in the original machine draft for the first model of the range.
2. To provide the gist of a foreign text quickly to those unfamiliar with the language. At the very least, the reader of the machine translation can decide which documents are relevant and therefore worthy of the expense of professional translation.
3. To provide the full vocabulary of words in a document. The reader might be fluent in the foreign language, but lack the vocabulary especially for a highly technical article.

Problems with ambiguity are comparatively unusual when translating word-for-word (as is usually possible) between modern 'rich' European languages. Moreover, when ambiguities do occur, they tend to involve alternative meanings for the same speech-type, such as a noun or a verb. It is rare to find a single word in a sentence that might be interpreted as a noun or as a verb (for example, 'the bat' and 'to bat') and, when such words do occur, it is normally very easy to distinguish noun and verb use by the preceding article or preposition. The result is that modern machine translators contain little coding to make distinctions between ambiguities, and can easily be upset when asked to choose between 'not-left' and 'not-wrong' when translating the word 'right' from English. However, a Latin translator has to pay close attention to such ambiguities. A single Latin word, such as the common ***decreta***, may have twenty alternative translations (in this example; we have not attempted to discover a world record for the most alternative translations for a single Latin word). The 'rich' European languages rarely have more than one or two alternatives; frequently no alternatives at all.

Our own interest lies in the machine translation of Latin, an area that has attracted little attention from the commercial vendors of machine translators generally, and which appears to have proved excessively disheartening to the few academic specialists who have trod this road [3]. Most professional translators, especially of Latin, seem to be completely unaware of how large a part their wide general knowledge plays in the translation process. Teachers should also understand the problem, for it explains the difficulties experienced by students in the translation of Latin. Students tend to be young and are likely to lack, not only the Roman knowledge of their teacher, but also the teacher's wide knowledge of general affairs.

EXAMPLES OF AMBIGUITY

We wish here to point up a number of examples of 'real-life' Latin ambiguity, and why they are especially troublesome. Most are taken from translations made by our machine translator.

1. The modern translators [4] of the large inscription bearing the ***Res Gestae Divi Augusti*** have noted that the word ***quoque*** had two different meanings (***quōque*** = each; ***quŏque*** = likewise) depending on pronunciation. The translators cannot assign the correct meaning with certainty. Did Augustus intend to be ambiguous, or was the meaning evident to a contemporary reader from his wider knowledge of Roman affairs?

2. *Divina vindicta improbitatem eius vita adempta coercuit.* 'Divine vengeance curbed the wickedness of him with the life withdrawn'. Or perhaps: 'The life withdrawn curbed the wickedness of him with divine vengeance'. It requires actual general knowledge on the part of the reader to separate these translations, both of which are grammatically correct. A machine translator has no way of deciding between the two alternatives.

3. *Recusantem apud Aurelianum, qui tum imperabat, detulerunt orthodoxi...* 'Pleading before Aurelian, who then was ruling, the orthodox (Christians) have accused....'. *Recusantem* can be translated as 'rejecting' or, in law, as 'pleading', while detulerunt can mean 'have conveyed' or, in law, 'have accused'. It is evident to human readers, with their wide general knowledge, that 'rejecting' and 'conveying' are nonsense, while 'pleading' and 'accusing' are correct. Yet how is an automatic translator to know this? The phrase has occurred suddenly in an historical text and there are no obvious legal key words in the complete paragraph to alert the program that the subject matter has changed.

4. A particular problem with Latin occurs with the juxtaposition of two words (usually a noun and an adjective or a noun and a verb) where both words can be reversed in speech-type (e.g. the noun could also be a verb and the verb could also be a noun) and thus meaning. A startlingly common example occurs with the short phrase *liber primus*. This phrase appears at the head of a multi-volume text, and means 'the first book' (noun-adjective in Latin). However, it can also be translated as 'the free first-man (leader)' (adjective-noun). Why do we prefer the first translation? Because we happen to know that the next volume begins *liber secundus* - 'the second book'. But our translation requires actual knowledge that there exists more than one book, a fact that might not be apparent from even the closest study of the first book. In short, it is inconceivable that a machine translator could provide the correct translation for *liber primus* with any degree of certainty.
Other irritating cases of Latin reversible meanings appear with:
Aurelianus ortus parentibus modicis: 'Aurelian born from moderate parents'. Or perhaps 'Aurelian born with the obeying little amounts'?
amas reges: 'you love the kings' or 'you will rule the buckets'.
amas mensas: 'you love the tables' or 'the buckets measured'.
These reversible Latin pairs are depressingly common and require human general knowledge to separate correctly.

5. Similarly *ad fossam quamdam*. This ought to be translated as 'at a certain ditch'. A perfectly legal alternative translation is 'to the dug certain (thing)'. The over-loaded word *ad* is particularly exasperating, since it can be translated as 'to', 'at' or 'near'. There is a clear difference in English between 'to a ditch', 'at a ditch' and 'near a ditch', but the Romans did not distinguish. Perhaps the best catch-all translation for *ad* is 'towards'. *Fossam* is the noun for a ditch, but can also be derived from the past participle of the verb 'to dig', another indication of Latin's primitive origins where the same word and its inflection can double as a noun and as a verb.

6. The word *contentus* can be the past participle of two verbs (*contendo* and *contineo*) or could be one of two adjectives. *Rex est contentus*: the king is/eats stretched or exerted/held together/tense or strained/content. Which is correct? It is impossible to say without further knowledge of the text.

7. In Latin, virtually all adjectives and participles can be used substantively; that is, they can be used as nouns. Thus ***bonus*** can be translated as ‘good’ or as ‘good man’. Less frequently, a noun can be used as an adjective: ***gratia Domini Iesu*** - ‘with the grace of Lord Jesus’. A legal alternative translation would be ‘with gratitude of Lord to Jesus’.

8. ***Qui principi imperii Christianis clementem se praebuit.*** ‘Who to the prince of the command by (means of) Christians presented himself merciful’. Or is it ‘Who of the beginning of the command to Christians presented himself merciful’? There is no way of telling even from the complete text, although texts by other writers indicate that the second translation is correct. This ambiguity fooled even us.

9. A particular problem relates to a short Latin phrase sent by a user as an example of mis-translation by our translator: ***nulla impudica Lucretiae exemplo vivet.*** Lucretia has just been raped and commits suicide, explaining that ‘nobody unchaste with the example of Lucretia will live’. This is translated as ‘the unchaste nobody of Lucretia with the example will live’. Suppose, however, that we replace ***Lucretiae*** with ***Luceriae*** (meaning ‘of the town of Luceria’). Now the sentence ought be translated as ‘nobody unchaste of (from) Luceria with (this) example will live’. In short, changing from the proper name of a person to the proper name of a town has of itself drastically altered the word-order of the correct translation, and the complaint occurred only because the user knew the difference between a named person and a similarly-named town. The machine translation was not at fault, given its lack of the user’s knowledge. Note, too, in this example that in English the noun (nobody) unusually precedes its adjective, instead of following it.

10. Much Latin has survived over about 2,000 years in the form of inscriptions. These are fascinating in their own right, since they reveal the change in the Latin language over the centuries of empire (essentially no change until the barbarian invasions and provincial secessions of the 3rd century AD, loosening Latin spelling around the periphery of the empire). The absence of verbs in most inscriptions, however, renders the problem of ambiguous translation even harder for a machine translator, which has been taught to pick out ‘first the subject noun, then the verb’.

MEDIEVAL LATIN

After the barbarians had finally managed to wreck the Roman Empire in the 5th-6th centuries, the languages of the occupied former provinces diverged, completing the process begun in the 3rd century. The original widespread literacy became concentrated into far fewer hands, predominantly those of Christian monks and court scribes. Western Europe entered the medieval age, defined by Gibbon as the 1,000 year period between the fall of Rome in 476 and the fall of Constantinople in 1453 (we use a more pragmatic definition: from 600 AD, when most Latin dictionaries terminate, to 1500 AD after the Reformation or Enlightenment). In the absence of published dictionaries, the spelling of Latin started to drift badly. Many of the less-educated authors tended to spell-as-I-speak, a particularly severe problem if the writer had a pronounced local accent. As early as the mid 6th century, St Gregory of Tours, living among civil chaos in a Frank-controlled area of modern France, was mis-spelling classical Latin words like ***victoria*** as ***victuria***, changing the stressed vowel. St Gregory was not even consistent with his changes.

The spelling of Latin changed in a more general sense to emphasise consonants. Thus words classically spelled with an initial *imm-* become *inm-* and consonants become doubled or un-doubled. The letter 'h' drifts in and out of medieval spelling and the letter 'y' replaces, or is replaced by, the letter 'i'. For example, the word listed in Lewis and Short's Latin dictionary as *synemmenon* is found spelled as *synnemenon, synemenon, sinemenon, sinemmenon* and *synhemmenon.*

Medieval writers were also often ignorant of Latin grammar. The deponent verb is one that is passive in form but active in meaning, thus the classical verb:
abominor, abominari, abominatus, to deprecate or detest. *Abominatus sum*: 'I have detested'.
Some medieval writers invented an imaginary perfect form, *abominavi*: 'I have detested'.
Here is the ending of a longer medieval Latin sentence: *...ad iurem quod possim*.
The writer has forgotten, or never knew, that *ius, iuris* 'the Law' is a neuter noun whose object case is *ius,* and has invented *iurem* as the object. But *iurem* means 'I may swear' (*iuro, iurare*). This unexpected and unintended change from noun to verb leaves the preposition *ad* dangling and causes disaster when the machine translator attempts to order the words of the complete Latin sentence.

Another major problem for the reader of medieval Latin is the inconsistent invention of new words by unrelated medieval scribes. Latin was always amenable to the construction of compound words, so that (for example) *receptamentum* might be construed from *recepto* (take back, admit, harbour) and the suffix *ament,* loosely translating as 'act of take-back-ing, means of take-back-ing'. However, one writer uses *receptamentum* to mean 'harbouring of criminals', while another uses it to mean 'right of entertainment'. Some Anglo-Norman scribes created Latin words by the dubious expedient of attaching Latin inflections to their native Anglo-Norman words. As Latham [5] observes in the introduction to his medieval dictionary, such words must often have been incomprehensible to Latin readers in the next town, let alone to readers in Italy.

Medieval monks in mainland Europe invented new words to describe the theory and practice of music, largely for the singing of hymns. With these developments, the word *plaga* became useful as a term for a musical stroke, as of a plectrum on a string. Medieval writers then added the words *plagis* and *plagius* to describe certain types of musical chord. Since *plagis* shares some inflections with the older *plaga* (which in turn is an overloaded Latin stem with several alternative meanings), the ambiguity of *plaga* and *plagis* became increased.

Worse is the adaptation of existing words for new purposes. Christian terms superseded older pagan meanings for many religious words. There was already a Latin word *baro, baronis, masc*, with meaning 'blockhead'. Norman court officials chose to invent a new word, *baro, baronis, masc,* to describe their feudal over-lords, the barons. Thus important late medieval documents, such as the famous *Magna Carta* (1215), the pact between a Norman king and his barons, make frequent references to the king's loyal blockheads.

Another unnecessary change contributing further to Latin ambiguity was the medieval contraction of the common inflection *-ae* to *-e*. This results in instant confusion since, for example, the word *domine* might mean 'O master' or 'of/to the mistress'. Several of the inflections of *plaga* ending in *-ae* now became inflections in *-e*, and thus overlapped with the ablative case of the medieval chord name *plagis*. The contraction is not even used

consistently. Texts attributed to Albertanus of Brescia (13th century) have the contraction in some of his sermons, but not in others. The authorship of ancient texts has been disputed on flimsier grounds than this, but we do not intend to propose a new theory about the real writer of Albertanus' sermons.

Finally, the well-known fact that medieval writers thought nothing of deliberately distorting the truth, and fabricating or amending ancient texts, poses a problem to the human reader but not, fortunately, to the machine translator.

MODERN (NEO-)LATIN

After the medieval age, and with the coming of the Enlightenment (or Renaissance) in the 15th century, the widespread reproduction of Latin texts by printing and an increased desire to seek the truth in all matters resulted in the spelling of Latin reverting to the classical style. Moreover, Latin remained in use by writers, such as Newton, Gauss and Descartes, all the way up to the modern Vatican bureaucracy. Indeed, it has become necessary to add new Latin words to describe recent developments in science and technology (such as the car and the aeroplane), and thus adding further potential for ambiguity. The modern Latin word ***virum*** means 'virus', but its inflections overlap with those of the classical Latin word ***virus*** (venom). More generally, as with medieval Latin, modern meanings have been added to existing Latin words, making again worse the problem of ambiguity of Latin stems.

SOME SOLUTIONS

It is worth recapping the essential problems of Latin translation briefly here.
1. The stems often have many meanings.
2. The inflections often apply to several cases.
3. Therefore a large dictionary is required in order to pick between inflections for different words.
4. The word order is generally not the same as that of English.
5. Medieval scribes lacked knowledge of Latin grammar and spelling, and independently invented the same new compound Latin words with different meanings.
It is not, perhaps, surprising that there are so few machine translators of Latin to English!

Translation of Latin texts requires a sufficient dictionary. We have supplied one: currently of some 37,000 unique classical Latin stems. Our ability to understand medieval and modern ('neo-Latin') words has been improved by the incorporation of a further 4,000 of the most common medieval words cited in Latham's medieval dictionary, and the direct inclusion of 2,500 words from the 'Calepinus Novus' neo-Latin dictionary, courtesy of Guy Licoppe of the Belgian Melissa Foundation. Medieval mis-spellings are handled by a variety of pre-stored character substitution patterns, and the ability to match phonetic (mis-)spellings with a phonetic dictionary.

Any machine translator worthy of the name should make at least some effort to distinguish the grammatical use of a word that might be a noun or a verb (or other speech-type). Our translator does indeed provide powerful AI procedures (grammar matching and word scoring) in an effort to resolve such ambiguities. However, if these techniques prove to be ineffective, the translator assumes simply that the most frequently-cited meaning is the correct one to use for the current Latin word. This is an example of the application of Bayesian statistics (where

results are biased by observation or experience) to our translator, and the frequency citation must be reasonably accurate if the strategy is to be of value. The frequencies are taken from the appearances of Latin words in standard paper dictionaries, and sometimes have had to be modified in the light of experience. For example, the noun ***multa*** (‘penalty’) is widely cited as very frequent, but ***multa*** actually occurs far more often as the feminine form of the adjective ***multus*** (‘much’). The problem of ‘reversible words’ (where two adjacent Latin words may be translated as either a noun-adjective or adjective-noun pair) may similarly be ameliorated by selecting the combination most commonly cited.

We spent some effort trying to ‘remember’ information from one sentence to another, so that a 3rd person verb could be tied to its subject matter, and elementary books of Latin Grammar even state that there is a rule: ‘in Latin, the subject is continued from the previous sentence unless there is clear indication otherwise’ [6]. Unfortunately, the ‘subject’ is not necessarily the Latin word with the nominative case in the preceding sentence, and this rule is heuristically worthless for real Latin sentences, as we soon discovered. For example:
rex amat canem suum, qui habitat in stabulo. humat ossum.
‘The king loves his dog, who/which lives in a kennel. He/it buries a bone’.
Did the king or the dog bury the bone? It is not even clear who lives in the kennel. And ***humat ossum*** can be translated as ‘the bone buries’. Latin writers tend to assume that readers can work these matters out for themselves, but an automatic translator lacks the required general knowledge.

Some of the difficulties of over-loaded Latin words are addressed in the electronic dictionary. It is possible to trim the dictionary output (that is, to restrict the words available to the translator) on the basis of their ‘age’ (whether the word could have been used when the text was written) and of their ‘area’. The latter refers to the type of text in which the word is used, divided by categories of general, legal, ecclesiastical, military, poetic/dramatic, scientific or technological usage. Our machine translator will allow the user to select the ‘age’ and ‘area’ manually, or it can make its own decision on the basis of which ‘area’ predominates in recent sentences. By far the biggest weakness of the latter approach is that the original writer may have ***intended*** that a Latin word retain its general meaning within a specialist text for which the word has also a specialist meaning. Trimming by ‘age’ is more valuable since, for example, it enables overlapping inflections of the modern word ***virum*** (‘virus’) to be removed with certainty from ***virus*** (‘venom’) in classical texts.

Because the machine translator lacks general knowledge, we have sought to apply ‘context’ to some Latin words with a ‘neural network’. The principle of a neural network is to seek to find non-obvious patterns in apparently random data, by training of the network on a test series whose results have been assigned by a human professional. We have examined the ambiguous word ***plaga*** by this means. One in ten of all occurrences of ***plaga***, and its inflected forms, in our test set of 24 million Latin words (of all ages) were manually examined, and the meaning of ***plaga*** assigned in each sentence where it occurred. These sentences were used to train a neural network, so that the trained network can be used to apply ‘context’ to all the other occurrences of ***plaga***.

Hitherto, our machine translator had always assigned the meaning ‘blow’ to ***plaga***, since this is most often the correct translation. One is reminded of the stopped clock, which is more accurate - twice a day - than the clock that loses just one second per day. However, this is not

really a satisfactory solution. The trained neural network assigns the correct meaning to ***plaga*** in about two-thirds of sentences examined. This is much better than the original invariant process, but the neural network never finds 'curtain' as the meaning - there are simply too few examples of use for training purposes. However, for the same reason, this cannot be a serious problem.

The neural network provides some fascinating insights. Consider the sentence ***rex amat plagam*** (a near-meaningless phrase, chosen especially so that readers cannot apply their general knowledge to provide the translation). What does ***plagam*** mean here? The professional translator would demand to see more of the text before attempting an answer. Our machine translator unhesitatingly provides the translation: 'the king loves the net (trap)'.

Other words examined and trained for use in the neural network include ***contentus, liber*** (un-inflected), ***saltus, -us***, and ***lustrum***. The effort required to train a neural network is considerable, and there is little chance that it will ever be available as a user option. At present, users are being asked to state whether they find the contribution of the neural network beneficial. If so, more examples will be added in due course. However, neural networks can never be a satisfactory substitute for general knowledge. The network makes the unfounded supposition that all writers of Latin deploy the same meaning of an ambiguous word in the same way, and is unduly biased by the historical pattern of usage of a Latin word. Modern Latin writers might completely reverse the historical usage. Perhaps all future references to ***liber primus*** really will mean 'the free first-man', instead of 'the first book'.

A very ambiguous language is always likely to be formula-driven; that is, the native speakers would use standard formulae to express certain phrases. Our translator allows many double-word phrases to be translated (such as the various forms of ***res publica***, as 'State'), while users may add their own phrases to a maximum of 100 clauses each of up to five Latin words.

Modification of our translator so that it would pick up the medieval inflection ***–e*** as a feminine noun/adjective alternative to ***–ae*** resulted unfortunately in a small number of masculine proper names of the 2nd declension (vocative case) also being attributed an ***–ae*** inflection and made thereby into feminine proper names. Considerable new code was required to resolve this difficulty.

Word ordering has been handled by the SVOE (subject-verb-object-else) routines, that move identified blocks of Latin text around into correct English word order. The big problem here is that a mis-assignment of an ambiguous Latin word (eg ***reges*** as a noun when a verb was intended) can cause a catastrophic misarrangement of word order. This propensity is to some extent curbed by use of a 'complexity algorithm', whose purpose is to decide whether the text should be re-ordered.

CONCLUSIONS

Our translator provides a quality of translation sufficient for the gist of a Latin text to be grasped, and to provide a fast aid to professional (human) translators. The problem of grammatical, but inaccurate, translation of ambiguous Latin text has proved to provide by far the most common source of complaint about our translator. This is due to the translator's lack of general knowledge, and the failure of users to comprehend the extent to which they use their own general knowledge with Latin text.

The ambiguities of the Latin language are far greater than those experienced with modern west-European languages. We believe that it will not be possible to improve significantly on machine translations of Latin without a means of conferring ‘context’ on each word as it is translated. Our attempts to provide context have been described in this article.

REFERENCES

1. Lewis & Short’s Latin Dictionary, Oxford University Press (1879; 2002 reprint).
2. Columbia Electronic Encyclopaedia (6th Edition, 2000), ***The Latin Language***, Columbia University Press. [www.encyclopedia.com/articles/07250.html.]
3. For example the Praelector program, at www.collatinus.org (Académie de Poitiers, France). The group stopped development work on the translator on 20th February 2002.
4. P.A. Brunt and J.M. Moore (Eds.), ***Res Gestae Divi Augusti***, Oxford University Press (1967; 1989 reprint), pg 78.
5. R.E. Latham, ***Revised Medieval Latin Word-list***, Oxford University Press (1965).
6. Collins Latin Dictionary Plus Grammar, page 134, Harper-Collins (1997).

ABOUT THE AUTHORS

William Whitaker and John White are the co-authors of the machine translator ‘Blitz Latin’. A free trial version of the translator can be downloaded from the independent distributors at www.software-partners.co.uk.

APPENDIX

This appendix contains more information for users of Blitz Latin, and was not part of the original paper.

MEDIEVAL LATIN CAPABILITY ENHANCED

We have enjoyed quite a lot of feedback from users, the majority of whom seem to be experienced with Latin but find it a very useful tool for interpreting predominantly medieval texts. For this reason, the medieval dictionaries have been sharply increased, and registered users receive a free 4,000-word dictionary of the most common medieval terms. This, coupled with the program's abilities to check phonetic mis-spellings against the whole dictionary and to create compound Latin words, ensures that the great majority of medieval Latin words can be translated. In addition, rare new medieval Latin words are added to the basic electronic dictionary as encountered: a prolific recent source was the Latin of Anglo-Saxon Charters from the 9th-11th centuries.

There remains, however, the problem of dismal spelling standards in the days before printed dictionaries. Should a modern dictionary of English attempt to include every child's mis-spelling of every word? We think not (neither do the publishers of such dictionaries), and therefore a certain amount of manual spelling correction is always likely to be necessary for medieval texts.

MEDIEVAL EFFECT ON THE SVOE ROUTINES

The writers of classical Latin chose to place words in order of their emphasis. However, this was probably never very popular in (spoken) Vulgar Latin, and the descendent languages of Latin rapidly adopted the convenient subject-verb-object-else (SVOE) word order that we see in modern west-European languages today. Medieval Latin underwent the same change. The SVOE word-ordering routines of Blitz Latin are technically very complex yet usually effective in placing classical Latin into modern word order but, as indicated in the main paper, a mis-assignment of an ambiguous word can cause a catastrophic breakdown of correct word ordering.

Since medieval Latin rarely requires this kind of word ordering, disasters caused by ambiguity can be simply avoided by using another option, entitled 'Verb after Nom(inative)'. Essentially this system just places the verb after any nominative noun, pronoun or adjective, without moving whole chunks of the sentence around. It is now the recommended solution for users attempting to translate late-medieval (say 9th century onwards) or modern texts. Apparently many users of Blitz Latin had discovered this preferred treatment for themselves.

THE ROMANS HAD DIFFERENT MENTAL IMAGES

We become used to images in our minds created by common usage or by various sayings that have stood the test of time. However, the Romans had different sayings and different mental images. This means that Blitz Latin may have translated a sentence accurately, but still it means nothing to the reader.

An obvious example is the way in which different peoples have described death. As well as the direct description, we may use phrases such as ‘passed away’, that would surely puzzle an ancient Roman. The Romans preferred to use ‘occisus’ (felled) and ‘sublatus’ (lifted) among other phrases. It is therefore necessary to use a little imagination in trying to comprehend a Latin text after translation by the literal-minded Blitz Latin.