

Stylistic Analysis Using Machine Translation as a Tool

Tess Crosbie
University of Bedfordshire

Tim French
University of Bedfordshire

Marc Conrad
University of Bedfordshire

Abstract

This pilot study investigates the potential of automatic translation engines to be used as a tool for literary analysis. Two sample texts, one prose and one poetry, were chosen as representative of their genre and translated from English into one of 62 languages. Re-translated back into English, the resulting texts were compared with a stylistic analysis of the original. Many stylistic features were retained although more subtle features of the texts were lost in the translation process. The similarity index varied widely between languages – between 90% and 32.9% -with corresponding stylistic feature retention observed. The similarity index gives a good indication of the literary quality of the text, suggesting that this process may be useful as a preliminary “filtering” technique in deeper studies.

1. Introduction

Automatic translation engines work by using leveraging probabilistic statistical machine techniques to generate a translation [1] from a source to a target language. This is further optimised via reference to an existing corpus based on the assumption that the desired translation already exists [2]. Even though an exact match may not be available, linguistic patterns (as embodied in speech or the written word) can be analysed to produce a convincing and stylistically appropriate translation. We go on to show that it is possible to preserve many stylistic features when artificially translating a text into first, a given target language and second, retranslating back into English. The results, though necessarily confined to fine grained stylistic elements such as metaphor, simile and related textual features, indicate that a high degree of stylistic preservation can be achieved during this process. Whilst the degree of similarity achieved varied greatly between 62 target languages (with a maximum similarity of 90% and a minimum of 32.9%) it is clear that some fine grained stylistic features did not survive the double translation process. Indeed, we postulate that the more subtle and global stylistic features, including cultural context and coarse-grained stylistic aspects (narrative patterns, hermeneutics) account for the sub-maximal

similarity scores achieved. Poetry, with its dependence on structure, proved particularly vulnerable although key features were retained in many translations. This has deep implications for extending our pilot study so as to examine longer texts of world class repute, hence in developing a set of computer gathered stylistic metrics that seek to reflect humanistic aesthetic, genre and cultural influences that any text inherently encompasses. Metaphor, as “a translation of unsayable experience [3]”, is particularly prone to mis-translation, as are idioms. This study investigates whether comparing the double translated result with the original text can act as an indicator of the extent to which stylistic properties are retained and so, by extension, provide a greatly simplified method of qualitative analysis.

2. Literature review

The danger of a human translator imposing their own personal bias on a given translation is inherent in the notion of translation itself. Many translators rely too heavily on the work of previous translators when working on a well known corpus, or they focus on delivering a version that seeks to please a given readership or otherwise try to align with editorial policy or external agendas (such as post-colonialism, feminism etc) rather than being true to the original text [4]. Thus, conscious or sub-conscious bias is likely to be present, leading to inaccuracies or stylistic ambiguities. However, machine translation is not simply a matter of replacing the human with the computer. Some languages, including English, use definite and indefinite articles but others, Latin and Russian amongst them, do not. This leads to particular problems when translating into English [5]. If the computer cannot differentiate between “an end” and “The End”, then metaphors and idioms are going to present a far bigger potential challenge.

Theoretically, this may not be a problem. Umberto Eco claims text is “a lazy machine...a device conceived in order to elicit interpretations” [6]. Taking this view, the computer could do a word for word translation and leave it to the reader to muddle through the best they can. In reality, however, greater accuracy is desired.

Samaniego Fernández [7] presents the following simple taxonomy of metaphor translation as follows:

- i. Metaphors cannot be translated. This view is substantiated by Nida (1964), Vinay and Darbelnet (1958) and Dagut (1976, 1987)
- ii. Metaphors are fully translatable. Kloepper (1981), Mason (1982) and Kurth (1995) are cited as evidence
- iii. Metaphors can be translated but with greater or less accuracy. The weight of evidence is on this categorisation, corroborated by Van den Broeck (1981), Rabadán Álvarez (1991), Toury (1985, 1995), Newmark (1980, 1988ab), Snell-Hornby (1988), Riedemann and Diéguez (1999), Schäffner (1997, 1998, 2004) and Ali (2006)

Previous studies into machine translation have tended to focus on speed rather than accuracy [8], [9]; however, there has been little attempt at qualitative analysis to determine the efficacy of the output. One exception is a semantic-based approach [10] for reducing the potential errors of English-> Arabic translations. This framework separates the process into two parts; a source interpreter parses, sifts through its list of ambiguous words and chooses the best version before determining the correct semantic. Then the target generator can select the appropriate word in its target language and construct the sentence and assign the correct tense. Similarly, a model integrating six types of semantic relationships between subject heads and then incorporating a SVM (Support Vector Machine) classifier to determine metaphor has been proposed [11], achieving F-1 in 89% of the test set for recognising metaphor. Another approach has been to use Gaussian mixture models to determine whether a phrase is being used figuratively [12]. This enables differentiation between phrases that are wholly figurative and those that are sometimes literal, producing up to 78% accuracy.

One of the few attempts to determine literary merit has been a stylistic analysis of Conrad's "Heart of Darkness" [13]. This work addresses what Stubbs calls "the Fish Fork", a logical trap observed by Fish in 1996 that stylisticians either describe only known linguistic factors and ignore anything outside these or describe only features already known to be important. This work concludes that computer analysis can offer a range of techniques for stylistic investigation and that the Fish Fork "applies to any study of anything [13]" and need not, therefore, be a barrier to this form of feature categorisation. Furthermore, the machine analysis actually revealed features not previously documented by critics.

This does not detract from the Chomskian view that literature must be taken in its context; however, Eagleton has observed that context changes, "Some texts are born literary, some achieve literariness, and some have literariness thrust upon them [14]". Theoretically, a re-translated text may become more

literary than its progenitor.

3. Method

Sample texts were taken from both prose – Hardy's "Far From the Madding Crowd" (Text A) and poetry – Shakespeare's Sonnet 18 (Text B). The choice of texts was determined by several factors. Both texts are relatively well-known so there was no difficulty in finding plenty of existing literary criticism. Text A contains a wealth of literary devices (alliteration, onomatopoeia, symbolism) and is a key point of the novel. Text B is one of Shakespeare's most popular sonnets and as well as being an excellent example of the genre, has a strong structure with the use of anaphora and personification.

Due to the limitations of the online tools, these were necessarily short texts (under 100 words). The texts were translated using automatic translation then the results were translated back into English. While different free online translation tools were tried, including Yahoo's BabelFish and the online version of Babylon, the most accurate results were produced by Google Translate. Several text comparison tools were tested and considered: KDiff, WinMerge, WordCompare and the free online version of Compare Suite. KDiff and WinMerge, both designed for programme comparison, tended to incorporate discrepancies once a major difference was noted. Although WordCompare worked well, Compare Suite provided a graphic representation of the differences and was therefore a quick visual evaluation tool, plus providing a keyword similarity tool. However, it did restrict the text lengths to about one paragraph. Literary analysis of both texts was performed by human volunteers and the results compared to the re-translated texts to determine how many literary features remained.

4. Text analysis

Text A

"In an instant the atmosphere was transformed to Bathsheba's eyes. Beams of light caught from the low sun's rays, above, around, in front of her, well-nigh shut out earth and heaven--all emitted in the marvellous evolutions of Troy's reflecting blade, which seemed everywhere at once, and yet nowhere specially. These circling gleams were accompanied by a keen rush that was almost a whistling--also springing from all sides of her at once. In short, she was enclosed in a firmament of light, and of sharp hisses, resembling a sky-full of meteors close at hand. [15]"

Features that emphasise speed and movement.

The double alliteration of “In an instant” places great emphasis on the word “instant” so the reader is made aware of the speed of the change. The juxtaposition of “at once. In short” reinforces the suddenness of the transformation, the more so because “at once” is repeated in this short passage. Movement is suggested by the asyndeton of “above, around, in front” and by the paradox of “everywhere at once, and yet nowhere”. The word “springing” also emphasises movement.

Features that emphasise light and sound. “Beams of light” literally and metaphorically mirrors “sun’s rays” and in their respective positions either side of the caesura, emphasises the image of light while the reference to “earth and heaven” demonstrates the all-encompassing quality of the light. Nouns referring to the light are compounded by adjectives; “reflecting blade”, “circling gleams” and these are added to as the sound is introduced; “keen rush”, “sharp hisses”. The author’s intention is to make this a sensuous description. The consonance of the “st” in “almost a whistling” coupled with the alliterative “s” of “sharp hisses” creates an onomatopoeic effect, creating a sound for the “meteors”.

Features that emphasise sex. Using “the atmosphere” as the subject of the sentence accentuates its importance in the subsequent passage and reiterates Bathsheba’s position as bystander. The imagery of Troy’s blade is phallic, particularly as Bathsheba is behaving improperly by being alone with Troy, and the chapter title (although not included in this extract) is called “The Hollow Amid the Fens”, suggesting a secret hiding place where a seduction could take place as well as echoing Bathsheba’s sexual “hollow”.

Being “enclosed in a firmament of light” is a reference to Genesis and the Creation, again trapping Bathsheba in a metaphor of sexuality and fertility. The simile is left to the end, “resembling a sky-full of meteors” to leave the reader with that vivid image in his/her mind.

Text B

“Shall I compare thee to a summer's day? Thou art more lovely and more temperate. Rough winds do shake the darling buds of May, And summer's lease hath all too short a date. Sometime too hot the eye of heaven shines, And often is his gold complexion dimmed; And every fair from fair sometime declines, By chance, or nature's changing course untrimmed. But thy eternal summer shall not fade Nor lose possession of that fair thou ow'st; Nor shall death brag thou wand'rest in his shade, When in eternal lines to time thou grow'st. So long as men can breathe or eyes can see, So long lives this, and this gives life to thee. [16]”

Structure. This sonnet is an example of typical Shakespearean style, comprising three quatrains in iambic pentameter ending in a heroic couplet, following a rhyming scheme of abab cdcd efef gg. It follows the tradition of dividing the sonnet into two parts. In the octave, Time is shown as the enemy of the transitory nature of beauty and there are references to different passages of time, “day”, “May”, “date”, “summer”. After the volta, highlighted by “But”, the sestet introduces Time as the solution: the youth’s beauty will be everlasting as long as the sonnet exists and the references are to the “eternal” and “So long as”. The final couplet, although part of the sestet, could stand alone and provides a strong closing point.

Technical devices. It is significant that there is only one enjambment – every line except line 9 finishes with punctuation. This is a poem of stated facts rather than rambling musings. Repetition (“more lovely and more temperate”, “every fair from fair”) and anaphora (lines 6 and 7, lines 10 and 11, lines 13 and 14) are used heavily throughout the sonnet. These techniques are used for emphasis, to accentuate the point being made. Contrasts are emphasised by antithesis, “more temperate./Rough winds” and the last word of lines 5 and 6, opposing “shines” with “dimmed”. Alliteration, a linking device, is lightly used which makes it more effective when it does appear, “chance, or nature’s changing course”, used at the end of the octave. The next use is in the final line, “long lives this, and this gives life to thee” where the double alliteration of the “l” and “t” force the line into prominence.

The object of the sonnet. The poem begins with a rhetorical question to “thee” (commonly assumed to be a youth) [17] so it seems as though the poem is going to be about the young man. However, the stressed “I” of the first line contrasts with the unstressed “Thou” of the second, foreshadowing the theme of the poem; it is less a tribute to the youth’s beauty than a proclamation of the writer’s skill and his assurance that his poem will be a future classic. This suggestion is furthered in the 12th line, “in eternal lines”, referring to the lines of the poem. Shakespeare has broken the fourth wall by acknowledging the poem and the existence of readers.

Use of Metaphor. Personification occurs throughout the poem in the form of Summer (“summer’s lease”), the Sun (“his gold complexion”), Nature (“nature’s changing course”) and Death (“shall death brag”). Summer and Death are personified to suggest a human relationship: Death is a rival for the poet’s love. The “summer’s lease” is echoed in line 8 at “thou ow’st”, extending the metaphor further.

From the first line, Shakespeare invites a comparison with summer and this continues through to the final couplet. Summer, generally presented as the perfect season, falls short of the youth's perfection and is unworthy to be compared to him. Summer has "Rough winds", and "too short a lease" while the youth's "eternal summer" is reinforced at the beginning of the sestet.

The sun is represented as "the eye of heaven". The "gold complexion dimmed" can be interpreted both as the sun's strength and beauty tarnished by clouds, just as the youth's beauty will be tarnished by time, but also "complexion" can be read as "temperament" (i.e. a combination of the four humours). This latter interpretation echoes "temperate" of line 2 effectively. [18]

Nature's "untrimmed" has a double meaning. It can mean either unadorned, indicating that Nature will strip the youth of his "fair" beauty but can also refer to the sails of a ship, suggesting that Nature's course is unadjusted. However, Jungman [19] has suggested that the "untrimmed" may actually mean "unadjusted" and therefore Shakespeare is saying that the thing that remains unchanged is Nature's changing, "mutability is eternal". This interpretation strengthens the structure of the sonnet with the octave representing change and the sestet reinforcing the endurance of the written words.

5. Results

Sixty-two different languages were used to translate the texts from English and back again. (Raw translation result data is available online at <http://research.tesscrosbie.com/index.php/8-translation/1-snippet-translation-results>)

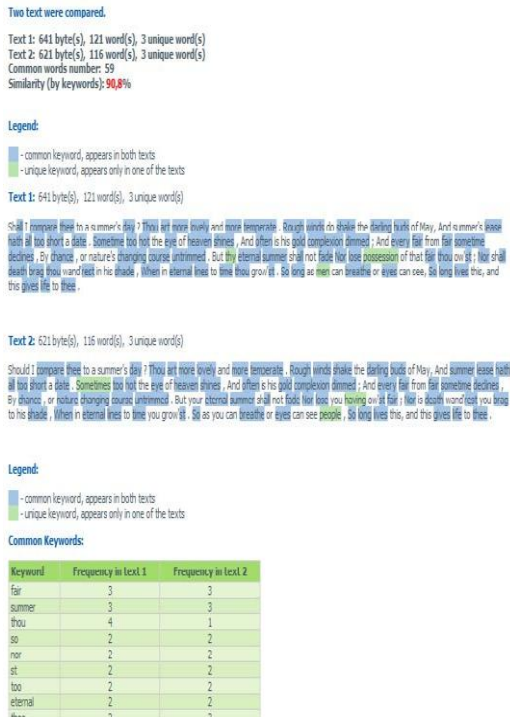


Figure 1. Results obtained for Text B: Filipino target language

Overall, the greatest similarity was achieved with the sonnet with an average of 72.85% of the original text remaining, compared to an average of only 55.3% for the prose. Filipino produced the most accurate results with 90.8% similarity between the two versions of Sonnet 18 (Figure 1). Catalan produced the least similar version with 32.9% similarity between the prose versions (Figure 2).



Figure 2. Results obtained for Text A: Catalan target language

Text A translation results

Dutch: 74.2% similarity to the original. In an instant the atmosphere was transformed into a Bathsheba eyes. Beams caught the rays of the low sun, above, around, for her, almost shut out earth and heaven -all emitted in the beautiful evolutions reflect Troy leaves, which seemed everywhere at once, yet nothing special. These circling gleams were accompanied by a keen rush that was almost a whistling -also come from all sides of it at once. In short, she was enclosed in a firmament of light, and of sharp hisses, resembling a sky full of meteors close at hand.

Latvian: 77% similarity to the original. In the immediate atmosphere was transformed into the eyes of Bathsheba. Beams of light caught from the low sun's rays, above around in front of her, well almost removed from the earth and heaven -all emitted in the wonderful events of Troy reflecting blade, which seemed to be everywhere at once, and yet nowhere specially. These circling gleams followed by a keen rush that was almost whistling -also springing from her hand all at once. In short, she was enclosed in a firmament of light, and sharp hisses, resembling a sky full of meteors close at hand.

Norwegian: 83.1% similarity to the original. In an instant the atmosphere was transformed to Bathsheba's eyes. Rays of light trapped by the low sun's rays, above, around, in front of her, well almost shut out earth and heaven -all emitted in the great evolutions of Troy's reflecting blade, which seemed everywhere at once, and yet nothing in particular. These circling gleams were accompanied by a keen rush that was almost a whistling -also springing from all sides of her once. In short, she was enclosed in a firmament of light, and sharp hisses, resembling a sky full of meteors close at hand.

Catalan: 32.9% similarity to the original. In an instant the atmosphere was transformed in the eyes of Bathsheba. The light rays trapped rays of the sun down, over, around, in front of her, almost excluding land and sky -all the wonderful changes in the cast sheet reflecting Troy, which seemed everywhere, and none in particular. These flashes were sometimes

accompanied by acute fever was almost a whistle are flowing around it immediately. In short, he was locked in a vault of light and sharp whistles, like a sky full of meteors in hand

Latin: 37.2% similarity to the original. In a moment the air was transformed to Bathsheba the eyes. Rays of light in front of him almost to the exclusion of taking a low rays of the sun between the earth and the sky above -all reflecting the emission of miracles in the course of Troy, the grass, which seemed everywhere at once, and yet never properly. These are the embrace he rushed shine with the keen hissing was near -and at the same time from all sides thereof. Finally that closed the firmament of light, and hisses like a sharp, the air full of meteors close at hand.

Table 1. Feature comparisons of Text A

Feature	Present in re-translated text Norwegian (83.1%)	Present in re-translated text Latvian (77%)	Present in re-translated text Dutch (72.4%)	Present in re-translated text Latin (37.2%)	Present in re-translated text Catalan (32.9%)
Alliteration "In an instant"	Yes	No	Yes	No	Yes
Juxtaposition of "at once. In short"	Yes	Yes	Yes	No	Partial
Repetition of "at once"	Yes	Yes	Yes	No	No
Asyndeton of "above, around, in front"	Yes	Yes	Yes	No	Partial
Paradox of "everywhere at once, and yet nowhere"	No	Yes	No	No	Partial
"springing"	Yes	Yes	No	No	No
"Beams of light" mirroring "sun's rays"	No	Yes	Yes	Partial	Partial
"earth and heaven" expression	Yes	Yes	Yes	Partial	No
Adjective of "reflecting blade"	Yes	Yes	No	No	No
Adjective of "circling gleams"	Yes	Yes	Yes	No	No
Adjective of "keen rush"	Yes	Yes	Yes	No	No
Adjective of "sharp hisses"	Yes	Yes	Yes	No	No
Assonance of "almost a whistling"	Yes	Yes	Yes	No	Partial
Alliteration of "sharp hisses"	Yes	Yes	Yes	No	No
Onomatopoeic "st" and "s"	Yes	Yes	Yes	No	No
Subject of sentence "the atmosphere"	Yes	Implied	Yes	Yes	Yes
Phallic "blade"	Yes	Yes	No	No	No
Trapping of Bathsheba by "enclosed"	Yes	Yes	Yes	No	No
Expression from Genesis, "firmament of light"	Yes	Yes	Yes	Yes	No
Simile of "sky-full of meteors"	Yes	Yes	Yes	Implied	Yes

Text B translation results

Filipino: 90.8% similarity to the original

Should I compare thee to a summer's day?
Thou art more lovely and more temperate.
Rough winds shake the darling buds of May,
And summer lease hath all too short a date.

Sometimes too hot the eye of heaven shines, And often
is his gold complexion dimmed; And every fair from
fair sometime declines, By chance, or nature changing
course untrimmed.

But your eternal summer shall not fade Nor lose you
having ow'st fair; Nor is death wand'rest you brag to his
shade, When in eternal lines to time you grow'st.

So as you can breathe or eyes can see people, So long
lives this, and this gives life to thee.

Hindi: 89.2% similarity to the original

Shall I compare thee to a summer's day? Thou art more
lovely and more temperate. Rough winds do shake the
buds of May, Sir, And summer's lease up all too short a
date.

Sometime too hot the eye of heaven shines, And often is
his gold look dim; And every fair from fair sometime
declines, By chance, or nature's changing course
untrimmed.

But thy eternal summer shall not fade Nor lose
possession of that fair thou ow'st; Nor shall Death brag
thou wand'rest in the shade, When in eternal lines to
time thou grow'st.

So long as men can breathe or eyes can see, can So long
lives this, and this gives life to thee.

Macedonian: 87.7% similarity to the original

Shall I compare thee to a summer day? Thou art more
lovely and more temperate. Rough winds do shake the
darling buds of May The lease is too short summer date.

Sometime too hot the eye of heaven shines, And often is
his gold complexion dimmed; And every fair from fair
sometime declines, Coincidentally, or changing the
course of nature untrimmed.

But thy eternal summer shall not fade Do not lose
possession of that fair thou ow'st; Nor death I will praise
wand'rest in his shadow, When in eternal lines to time
thou grow'st.

So long as men can breathe or eyes can see, So long
lives this and this gives life to thee.

Korean: 53.8% similarity to the original

I compare the summer day you do?
You art more lovely and more temperate.
Rough winds do shake the bud a lovely February
Summer's lease is too short a date in the dark.

Sometimes it's too hot in the sky shining eyes,
And often his gold complexion dimmed;

When all the processes in the process, rejection
By chance, or nature's changing course untrimmed.

But thy eternal summer will not fade
Ow'st you have also lost equity;
In addition, death, to thee in the shade boasts
wand'rest
When you time a permanent line in. grow'st

This person can rest or one eye to see
Too long for this life, it gives life to thee.

Latin: 50% similarity to the original

If your summer compare this day
You are more handsome and more temperate.
Changes darling buds of May rough winds
And he has a short summer course too friendly.

Once too hot the eyes of heaven shines,
And often, the gold complexion dimmed;
Every fair is now on equal terms, retired,
It may be promised and changing the course or
nature.

Not disease but eternal
Nor lose possession of that fair thou ow'st;
Neither the death of you in the shadow of his more
wand'rest,
Then into the field with grow'st everlasting.

As long as they can breathe or eyes can see,
While life, and gives life to thee.

Table 2. Feature comparisons of Text B

Feature	Present in re-translated text Filipino (90.8%)	Present in re-translated text Hindi (89.2%)	Present in re-translated text Macedonian (87.7%)	Present in re-translated text Korean (53.8%)	Present in re-translated text Latin (50%)
Iambic pentameter	Yes	Yes	Yes	No	No
Rhyming scheme abab cdcd efef gg	abab cdcd efef hi	abcb dedf ghgh ij	abab cdcd efef hh	abcd efef hijk ll	abcd efef ijkl ll
Clear difference between octave and sestet	Yes	Yes	Yes	Yes	Partial
Time references in octave, "day", "May", "date", "summer"	Yes	Yes	Yes	Partial	Yes
"But" at volta	Yes	Yes	Yes	Yes	No
Expressions of endurance, "eternal", "So long as"	Partial	Yes	Yes	Partial	Yes
Strong final couplet	No	Partial	Yes	No	Partial
Little enjambment	Yes	Yes	Yes	Partial	Partial
Repetition in "more lovely and more temperate"	Yes	Yes	Yes	Yes	Yes
Repetition in "every fair from fair"	Yes	Yes	Yes	No	No
Anaphora of "And often...And every"	Yes	Yes	Yes	No	No
Anaphora of "Nor lose...Nor shall"	Yes	Yes	No	No	No
Anaphora of "So long...So long"	No	Yes	Yes	No	No
Antithesis of "more temperate. Rough winds"	Yes	Yes	Yes	Yes	No
Antithesis of "shines" and "dimmed" at end of lines	Yes	Yes	Yes	No	Yes
Alliteration of "chance, or nature's changing course"	Yes	Yes	Partial	Yes	No
Alliteration of "long lives this, and this gives life to thee"	Yes	Yes	Yes	Partial	Partial
Stressed on "I", unstressed on "Thou"	Yes	Yes	Yes	No	No
Broken fourth wall	Yes	Yes	Yes	No	No
Personification of summer	Yes	Yes	No	Yes	Partial
Personification of the sun	Yes	Yes	Yes	No	Yes
Personification of nature	Yes	Yes	Yes	Yes	No
Personification of death	Yes	Yes	Yes	Yes	Partial
Echoing of "summer's lease" and "thou ow'st"	No	Yes	Yes	Yes	No
Comparison with summer	Yes	Yes	Yes	Yes	No
Double meaning of "complexion"	Yes	No	Yes	Yes	Yes
Double meaning of "untrimmed"	Yes	Yes	Yes	Yes	No

Figure 3 shows a syntax tree of a small extract of Text A which lays out the parts of speech. Although the same tree can be drawn using the higher similarity re-translations, attempts to reproduce it in Latin, "These are the embrace he rushed shine with the keen hissing", or in Catalan, "These flashes were sometimes accompanied by acute fever was almost a whistle" are not possible. Meaning is virtually lost. Other languages produced unexpected results in the translation. Some were understandable, such as the Bulgarian translation of "close at hand" into "at your fingertips" while the German translation of "These circling gleams" into "Even mushrooms" was not. In Persian, "the atmosphere" became "Joe". Occasionally, all meaning broke down, as in the Korean which inserts "a sharp gyeuidoeotseupnida Ballroom" at the end.

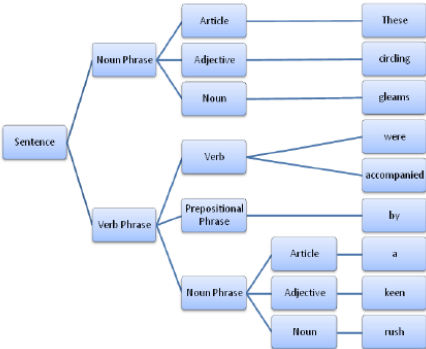


Figure 3. Syntax tree of extract

Text B produced fewer random elements into the re-translation than Text A but many Text B translations struggled with the auxiliary verb in “Shall I compare thee..?” with 42 of the 62 texts failing to translate it accurately. This phenomenon of inadequate linguistic synonyms has been documented by [20], highlighting the difficulties with ontologies that are unable to support multilinguality and observing the need for further work in this area. While initial observation suggests a high lexical similarity equates to a parity of experience, cultural and aesthetic differences as outlined in [20] mean this is not a guaranteed result. Even where there are no linguistic challenges, as in the line, “the darling buds of May” makes less sense to an antipodean who is accustomed to different seasons in May. The further apart in culture and shared history, the more exacerbated this result. Metaphors and idioms, although appearing to survive the double translation process may have not been understood in the target language. The “enclosed in a firmament of light” of Text A has connotations of Genesis, a subtlety possibly lost on members of non-Christian, non-Jewish religions. These are significant challenges.

6. Conclusion

This study is necessarily limited in scope and self-limited by the exemplar textual fragments chosen and by tools chosen to consideration of finegrained category and feature stylistic analysis rather than, for example, hermeneutics, narrative patterns and holistic deconstruction. A structuralist approach is only one way to approach literature and does not include the rich analysis provided by taking a poststructuralist, post-modern approach or using a feminist/Marxist/psychoanalytic critical view of the texts.

A high degree of similarity reveals that fine grained style (alliteration, use of adjectives, anaphora) is quite well preserved with viability induced by efficacy of black-box engines and prestored corpora. This may be one reason the results for ancient or “minority” languages (Latin, Catalan) show the greatest discrepancy. Overall, the results demonstrate the linguistic family etymology with greater similarities between Germanic than Romance languages. German itself, however, scores the same similarity in Text A as Basque, Chinese and

Vietnamese (51.4%) suggesting a specific issue with the German translation. However, more subtle aspects such as the use of similes and appropriate adjectives were not always well preserved. A maximum of 90% similarity between texts suggests literary excellence is reliant on implicit stylistic norms and cultural semantic contexts. These operate at aesthetic levels and will be the subject of our future work. Not surprisingly, there was no “infinite monkey” effect and none of the re-translated texts were an improvement on the original. None of the texts produced a literary device not seen in the original versions that could be interpreted as intentional. The “missing 10%” is therefore significant and the human experts regarded the missing components as being aesthetically detrimental as compared with original literary quality. In some instances, the text was virtually incomprehensible. The “missing 10%” can never be filled merely using the set

of metrics herein. The aim of our main study (this was a pilot) will be to identify, using two versions of a literary “masterwork”, the efficacy of using more subtle stylistics as well as fine-grained metrics so as to develop a test bed that can achieve a near human level of aesthetic literary stylistic discrimination, hence at least partially seek to find the “missing 10%”. We believe that it is within the power of the computer to deconstruct the more subtle stylistic attributes that are typically mediated by humans as determining the literary “merit” of a text. These attributes include use of “humour” as well as metaphor, simile and a wide range of similar literary stylistic constructs. This is our long-term vision.

7. Acknowledgements

With thanks to Helen Manley.

8. References

- [1] Google. www.translate.google.com/about. [Online]. <http://translate.google.com/about/index.html>
- [2] David Bellos, *Is That a Fish in Your Ear?: Particular*, 2011.
- [3] P. Vigdeman, "The Task of the Translator," *Kenyon Review*, p. 67, Fall, Vol. 26, issue 4 2004.
- [4] A. Paterson, "Translation As Editing," *World Literature Today*, Nov-Dec 2006.
- [5] Robert S Bolia and Raymond E Slyh, "Representation and Comprehension in Machine Translation and Intelligent Decision Support," *IEEE Intelligent Systems*, pp. 40-48, July/August 2011.
- [6] Umberto Eco, *Confessions of a Young Novelist.:* Harvard University Press, 2011.
- [7] Eva Samaniego Fernández, "Translation Studies and the Cognitive Theory of Metaphor," *Review of Cognitive Linguistics* 9:1, pp. 262-279, 2011.
- [8] Che-Yu Yang, "Cross-Language Instant Messaging with Automatic Translation," in *Fourth International Conference on Ubi-Media Computing*, 2011.
- [9] Aarne Ranta, Krasimir Angelov, and Thomas Hallgren, "Tools for Multilingual Grammar-Based Translation on the Web," in *Proceedings of the ACL 2010 System Demonstrations*, Uppsala, 2010, pp. 66-71.
- [10] S Jusoh and H.M. Alfawareh, "Automated Translation Machines: Challenges and a Proposed Solution," in *Second International Conference on Intelligent Systems, Modelling and Simulation*, 2011.
- [11] Xuri Tang, Weiguang Qu, Xiaohe Chen, and Shiwen Yu, "Automatic Metaphor Recognition Based on Semantic Relation Patterns," in *International Conference on Asian Language Processing*, 2010.

[12] L. Li and C. Sporleder, "Using Gaussian Mixture Models to Detect Figurative Language in Context," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, Los Angeles, 2010, pp. 297-300.

[13] Michael Stubbs, "Conrad in the Computer: Examples of Quantitative Stylistic Methods," *Language and Literature*, vol. 14, no. 1, pp. 5-24, 2005.

[14] Terry Eagleton, *Literary Theory: An Introduction*. Oxford: Blackwell Publishing, 2008.

[15] Thomas Hardy, *The Works of Thomas Hardy*.: Reprint Services Corporation, 2007, available from <http://books.google.com/books?id=VLabpq0vIYkC&printsec=frontcover&dq=bibliogroup:%22The+Works+of+Thomas+Hardy%22&hl=en&sa=X&ei=KAAOT5-IMqbm0QGMzcDzBQ&ved=0CEkQ6AEwBA#v=onepage&q=bibliogroup%3A%22The%20Works%20of%20Thomas%20Hardy%22&f=false>.

[16] William Shakespeare, "Sonnet 18," in *William Shakespeare: The Complete Works*, 1968th ed., Peter Alexander, Ed. London & Glasgow: Collins, p. 1311.

[17] *The Norton Anthology, The Middle Ages through the Restoration and the Eighteenth Century*, 8th ed., Stephen Greenblatt et al., Eds. New York: W. W. Norton & Company, 2006, vol. 1.

[18] Robert H. Ray, "Shakespeare's Sonnet 18," *Explicator*, vol. 53, no. 1, p. 10, Fall 1994.

[19] Robert E. Jungman, "Trimming Shakespeare's Sonnet 18," *ANQ*, vol. 16, no. 1, pp. 18-19, Winter 2003.

[20] Philipp Cimiano, Elena Montiel-Ponsoda, Paul Buitelaar, Mauricio Espinoza, and Asunción Gómez-Pérez, "A Note on Ontology Localization," *Applied Ontology*, pp. 127-137, April 2010.