A Hybrid Statistical-Linguistic Model of Style Shifting in Literary Translation

Meng Ji

Abstract [1] **.** The present paper presents an original inter-disciplinary study of style-shifting in literary translation, which draws upon methodologies and techniques from corpus stylistics and computational stylometry, and relevant sociolinguistic theories of style variation. Such an innovative approach to the literary translator's idiosyncratic use of language sets out to address one of the most difficult issues in textual stylistics, i.e. the cognitive rationale behind style-shifting in literary translation. Keywords: Textual statistics; Corpus stylistics; Computational stylometry; Style shifting; Multi-variant analyses; Context-motivated theory in literary translation; contrastive linguistics

1. Outline of the current study

It is argued in the present study that style-shifting in literary translation is a very complex phenomenon which requires a quantification of source text contextual information potentially explanatory to such an important creative process in literary translation, which has been rarely discussed in depth in past studies. To distinguish the cognitive nature of stylistic variation as a conscious strategy devised on the part of the translator from being a simple reflection of the translator's writing habit, the context-motivated theory (CMT) is formulated and put to test with primary linguistic data retrieved from a parallel corpus containing Cervantes's Don Quijote in seventeenth-century Castilian and its two modern versions in Mandarin Chinese.

The hypothesis-testing process has been greatly facilitated by the application and experimentation of statistical techniques which are widely used in social and behaviour sciences, despite that their productivity in text-based corpus stylistics remains largely under-explored. The interesting results obtained in the present study suggest that in the study of stylistic variation, which is a crucial representation of the creative nature of literary translation, an integral approach to the subject matter which combines the discriminating strength of quantitative statistical analysis with the explanatory power of adjacent sociolinguistic theories holds the key to a deeper understanding of the cognitive nature of the textual phenomena under investigation.

[1] Humanities, Imperial College London, SW7 2AZ
m.ji05@imperial.ac.uk

2. Context-motivated theory of style-shifting in literary translation

Style-shifting has always been closed associated with the study of context-motivated or proactive speech variation in sociolinguistics (see Labov, 1972; Bell, 2001; Eckert & Rickford, 2001), primarily in qualitative terms; while its exploration from a computational stylistic point of view seems to have been less discussed. However, the present paper will show that the topic of style-shifting may actually be further developed and better explained through the processing of quantitative linguistic data from purposely built corpora. From the outset, it should be pointed out that the current study differs essentially from a pure quantification of linguistic information in search of underlying patterns in translational texts; instead, it focuses specifically on the exploration of contextual factors or parameters which tend to characterize certain textual phenomena that have been highlighted in a previous statistical study as a result of their abnormal or unpredicted occurrence in the corpus texts (Ji, 2008).

In this sense, it may be said that the current paper distinguishes itself from many stylometry studies (Henderson, 1978; Hoover, 2001), for in the place of being concerned with the devising and improvement of techniques that may increase the sharpness or accuracy of authorship attribution implements, it aims to delve into the cognitive dimension of stylistic variation, i.e. context-motivated /background-foregrounding or habitual /unconscious, in literary translation.

The design of the current project is largely based on a hypothetical proposition which attempts to explain stylistic variation brought about by literary translators in their work, drawing on their indirect or assumed estimation of the contextual situation as depicted in the source text. It offers an alternative explanation to style-shifting and addresses the issue from a perspective that is somehow different from previous theories, such as Labov's attention-to-speech model (1972) or Bell's audience design (1984, 1997 & 2001). Its theoretical speculation is based on observations of a specific type of style-shifting, where the textual phenomenon is seen as a pragmatic strategy developed on the part of the translator.

While past studies have seen stylistic variation in literary translations as (H1) invariably source-text derivative and (H2) such dependence is supposed to be largely fixed at a linguistic level, it is argued here that holding the first hypothesis as true, H2 may not necessarily follow. That is because, in dealing with linguistically challenging or somehow intangible linguistic issues in literary translation, such as archaism in a typologically and diachronically distant language, more often than not, the translator may resort to contextual cues in the original text, such as audience, setting, the relative social status between the parties involved in a speech event, etc.,

rather than linguistic cues, in an effort to overcome the linguistic challenges in the historical source texts1.

A key set of terms lying at the heart of the proposed context-motivated theory (CMT) is the concept pair of dominant versus latent, which is put forward to describe contextual features that have either stimulative or alleviative effects on the translator's decision of using a certain linguistic device. For instance, in studying the use of archaism, it is easy to fellow that taking the surrounding audience as a variable of the contextual analysis, the presence of a listening audience in the source text may well trigger off a stimulative effect on the translator's mind for him to use an archaic idiom, which in turn will enhance the chivalry image of the enchanted knight. On the contrary, the absence of audience, or a private setting of communication tends to have an alleviative effect on the conscious-minded translator, for it is commonsense that in a private setting, it is unlikely for one to take on an ostensibly archaic tone for showing-off purpose. Following this line of argument, all the source text contextual features highlighted in the current study have been tagged manually as dominant or latent as an initial encoding for later statistical analyses.

Lastly, it is envisaged that in the final analysis, the either context-motivated or habitual nature of the stylistic variation detected in the literary translation will be sustained by two different outcomes of the statistical modelling. That is, on the one hand, if the statistical modelling of contextual factors as assisted by the multi-variant analysis technique has helped establish a statistical model in each subdivision in which the contextual factors as constituents of the constructed statistical dimensions are largely dominant, then we may argue that the detected style shifting has been a conscious decision made by the translator to explore an idiosyncratic profile of his own. On the other hand, if the result of the statistical test shows that there is a proved instability in the contextual features characterizing the style shifting, or in other words, the contextual features seem to linger between dominant or latent values, it may therefore be said that the stylistic variance is more likely to be due to an unconscious use of archaism by the translator or simply a reflection of his or her habitual writing habit.

3. Hypothesis testing

The multi-variant technique used is categorical principle component analysis, also known as CATPCA. It aims to build a statistical model with limited dimensions, usually two, on a wide range of variables. CATPCA may help arrange a large amount of original variables in a way that is susceptible to human observation in finding interesting patterns in quantitative textual analysis. This is a vital process in the exploration of primary textual data where the reduction of numerous variables into a limited set of statistical dimensions holds the key to a deeper understanding of the nature of the original dataset; and hence provides important clues to research questions raised around the fundamental structure of textual data as measured in a two or three dimensional space. Despite the wide use of CATPCA in social and behaviour sciences (Stevens, 1992; Thomas, 2004), the applicability of such modelling technique in corpus stylistics remains to be tested, in this sense, the research methods developed in the current paper will undertake some initial investigations into the potential productivity of CATPCA in corpus stylistics.

Of course, such a context-feature-motivated approach may only survive in the environment of literary translation, especially in translating historical texts; literary translation should be seen as a creative process in its own right, which thus allowing some flexibility and pragmatism in handling with certain issues that are hard to pin down.

In a previous study, it has been proposed that to facilitate a comparative study of the source text and its two modern Chinese versions, the first part of the Spanish novel, which contains fifty-two chapters, has been divided into ten thematic divisions. The use of multiple regression test has revealed that while in most subdivisions, there is a corroborated similarity between the two Chinese translations in terms of the occurrence of archaisms in the protagonist's speech, in the ninth subdivision, perceivable discrepancies seem to emerge, which then brings about the issue of style-shifting in Liu's work.

In line with the CMT, a highly desirable model of contextual patterns which may be applied to explain a translator's deliberate use of language consists in an apparent consistency of the occurrence of situational parameters with dominant values as opposed to latent values. With regard to the use of archaism, it may be argued that archaism as an important form of rhetorical device tends to be used in a formal rather than informal setting of communication, where the degree of formality of communicative settings may be reasonably quantified by various situational parameters assigned with their corresponding values. For example, the quantification of interrelation-centric factors such as formality of relationship (FR) between addressor and addressees may be divided into latent values and dominant values within the context of the current study on archaism. To be specific, the dominant value of FR may refer to a formal relationship between the protagonist and his listeners, while the latent value of the same factor may be construed as a rather informal relationship between the communicating parties.

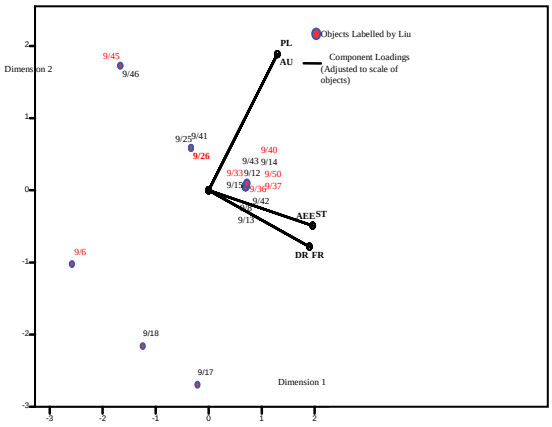
Table I Summary of VAF of Subdivision IX (Liu)

Total (Vector Coordinates)			
Dimension			Mean
1		2	
AU	0.319	**0.677**	0.996
ST	**0.732**	0.046	0.778
DR	**0.687**	0.116	0.802
FR	**0.687**	0.116	0.802
PL	0.319	**0.677**	0.996
AEE	**0.732**	0.046	0.778
Active Total	3.477	1.677	5.153
% of Variance	**57.948**	**27.943**	**85.891**

N.B. AU= Audience; ST= Social Status; DR=Distance of relationship; FR= Formality of relationship; PL= Place of communication; AEE= Addressee's evaluation of the content of speech;

Table I presents a summary of the two dimensional CATPCA model built upon the data retrieved from Liu's translation of Subdivision IX. As may be seen, the first dimension of the statistical model is primarily defined by the four interrelation-focused factors which are the distance of relationship (DR), formality of relationship (FR), relative social status of participants (ST) and finally, addressees' evaluation of the speech content (AEE). Meanwhile, the second dimension of the model is reduced to the two variables quantifying the communicative environment in which the protagonist has chosen to use archaic speeches or otherwise: audience and place of communication. As shown in the corresponding object-component-loading biplot of the statistical model, important patterns bearing on the distributional characteristics of IARS2 in relation to those of LTAS3 in Subdivision IX seem to emerge.

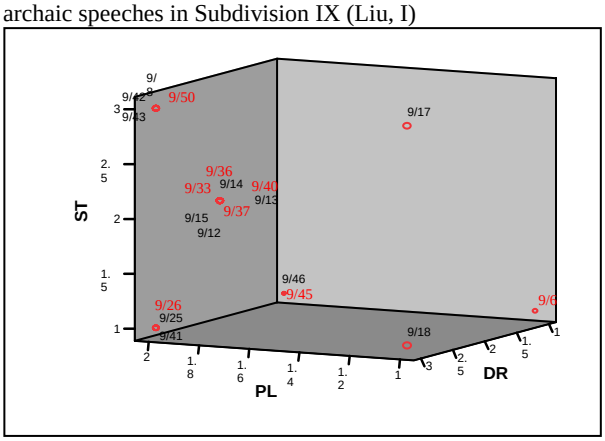
Diagram I Biplot of Subdivision IX (Liu, I)



Most object points in Subdivision IX, including IARS and LTAS, are clustered in the triangle space delineated by the two sets of variables functioning as the first and second dimension of the model. The clustering position is characterized by its restricted projection on the second dimension and its lower-intermediate value reported on the first dimension. To facilitate the discrimination of
IARS from LTAS, the former has been marked in bold, which as the graph suggests, is nested intensively within the area demarcated by its LTAS counterpart and is thus quite difficult to separate them apart. There are some outliers running out of the concentration area and distributed sparsely in the biplot, which in the case of IARS are 9/6, 9/26 and 9/45.

There are some outliers running out of the concentration area and distributed sparsely in the biplot, which in the case of IARS are 9/6, 9/26 and 9/45. However, an important commonality shared by these IARS outliers is that they all feature to the left of the origin, while the vectors invariably extend into the areas to the right of the centroid. This graphical quality of IARS outliers implies that their values on the factors quantifying the first dimension of the model, i.e. DR/FR and ST/AEE, may well be equally as low as those IARS clustered to the central-right of the graph.

Diagram II 3-D plot of the distribution of Don Quijote's archaic speeches in Subdivision IX (Liu, I)



In view of the inherent limitation, of a two-dimensional graph, which may cause the visual overlapping of actually quite distinctively distributed factors, we proceed to convert the biplot into a three-dimensional graph. It should be noted that such procedure does not entail a structural modification of the already set-up statistical model; rather, it serves as a kind of assistant visualizing tool which will allow us a more intuitive access to the complex data structure under investigation. The visualization function used is readily provided in the SPSS 15.0 version known as chart-builder. As part of the software requirements, only one variable has to be specified in each of the three axes on the spatial representation of the data structure.

In Diagram II, we may see that the tri-partite pattern of the distribution of situational parameters already seem to emerge, and the high uniformity in the quantification of variables sustaining each "branch" of the pattern greatly facilitates the construction of the three-dimensional graph. That is, we need no more than to select arbitrally one variable from each parameter set and subsequently fill it in the mould furnished by the chart-builder of the software. The result of the construction of the three-dimensional plot is shown in Diagram II. An important feature of Diagram II is that it clearly sets apart the virtually overlapped two clusters of objects on the biplot, which has been made possible through the erection of the third dimension and the following segregation of the two minimally discerned variable sets on the biplot, i.e. FR/DR, PL/AU. In fact, as shown in the 3-D graph, the super-cluster which appears to the central-right on the biplot is actually composed by two quite distinct clusters as evidenced by their different locations along the ST and DR scales on the 3-D plot. In terms of the distribution of IARS, which constitutes our main concern in the establishment of contextual patterns that may explain Liu's stylistic use of archaisms in his translation of Don Quijote's speeches in Subdivision IX, we can see that most of the highlighted scatters, i.e. IARS, may be covered by the X-Y plane on the spatial representation of the data configuration. The only exception is 9/6 and 9/18, whose high value on the z-scale implicitly requires the specification of the variable PL.

In spite of the fact that IARS in Subdivision IX seems to cover both the lowest and highest ends of the X and Y scales on the 3-D chart, it is also obvious that IARS in Subdivision IX mainly appear in a mixed setting of communication as quantified by their medium value on both the ST and DR

dimensions of the graph. The central cluster, which is composed of the following items 9/12, 9/13, 9/14, 9/15, 9/33, 9/36, 9/37, 9/40 is indicative of a contextual environment where the relative status and the interpersonal relationship among the parties to the exchange tends to be underspecified. The graphical analysis fits well into the general description of the textual plot, which with the exception of chapters when the protagonist is absent from the narrative scene, has shifted from the knight-squire private communication setting in the previous subdivision to the public domain as provided by the inn, which is also the title of the Subdivision IX.

4. Conclusion

Table II Summary of the CMT analysis

Statistical Dimensions	Value	Stylistic significance
ST (AEE)	Medium to high	Dominant
DR (FR)	Medium to high	Dominant
PL (AU)	High	Dominant

Table II presents the results of the CMT analysis of style-shifting in Liu's translation. In quantifying such narrative background with a view to establishing the patterns of contextual features that seem to characterize the use of IARS in Subdivision IX, we can see that the four interrelationship-focused factors come to the fore, i.e. DR/FR, ST/AEE. From here, it may be inferred that in his increased use of archaisms in don Quijote's utterance as detected in Subdivision IX, Liu seems to be rather sensitive to the interrelationship between the protagonist and his addressees; the contextual factor pair which focuses on the speech environment, i.e. PL and AU, also shows an unambiguous high value. As a result, it may be argued that according to the CMT, style-shifting, which has been detected in Liu's translation of Subdivision IX, is very likely to have been a conscious decision made on the part of the Chinese translator. The interesting finding uncovered through the formulation and testing of the context-motivated theory has thus provided us valuable and plausible explanations into the rationale or cognitive mechanism behind complex textual phenomena such as style-shifting.

Reference

[1] Bell, A. (2001) "Back in style: Reworking audience design", in P.Eckert and J. Rickford (eds.) Style and Sociolinguistic Variation, Cambridge University Press

[2] Eckert, P & Rickford, J (2001) Style and sociolinguistic variation, Cambridge University Press

[3] Labov, W (1972) Sociolinguistic Patterns, University of Pennsylvania Press

[4] Stevens, J (1992) Applied Multivariate Statistics for the Social Sciences (second edition), Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

[5] Thomas, B (2004) Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications, Washington DC: Ameican Psychological Association

[6] Ji, M (2008) "Quantifying style in two modern Chinese versions of Don Quijote", Meta: Journal des traducteurs, 53 (4), Quebec: Les Presses de l'Université de Montréal

[7] Yang, Jiang (1978) Tang Ji He De (Don Quijote), Beijing: People's Literature Publisher

[8] Liu, Jingsheng (1995) Tang Ji He De (Don Quijote), Gui Lin: Li River Publisher

[9] Henderson, Michael M.T. (1978) "Stylistic Variation and Underlying Structure", in Journal of Linguistics, vol. 14, no.2, 179-82

[10] Hoover, D (2001) "Statistical stylistics and author contribution: an empirical investigation", in Literary and Linguistic Computing, 16 (4), 421-44

[11] Cervantes, Miguel (1605) Don Quijote de La Mancha (part I), Francisco Rico (ed.), Instituto Cervantes online version http:// vc.cervantes.es/obref/quijote/