

# Correspondence analysis of international relative deviance

PETER JOHN HASSALL\* and SIVA GANESH\*\*

**ABSTRACT:** The synchronic etic approach outlined in this paper is designed to stimulate interest in the study of English as an international and intranational language, consistent with the aims of the International Association of World Englishes (IAWE). It suggests that certain aspects of different World Englishes (WE) may be compared to each other by considering those surface linguistic features that are shared in response to a given language task and reports the results of research designed to analyse international relative deviance (relDEV.EIL) between a number of world Englishes. The study employs Correspondence Analysis (CA) to compare electronic corpora compiled from the written English of groups of tertiary students in different countries whilst undertaking an identical language task. The WE corpora to be compared were assembled from data<sup>1</sup> gathered in tertiary institutions in Japan, South Korea, Taiwan, Thailand and Ras Al Khaimah in the United Arab Emirates. In the present study, the Word frequency program WORD developed by Nation et al. (1988) was utilised together with the SAS System to provide a graphical representation of CA. An analysis of relDEV.EIL was compiled of the orthographic forms selected by the WE user groups in the five countries in response to the language task. This enabled a more complete picture to be built up of just how the groups differed from each other with respect to the orthographic forms they used. Implications for English language teaching are discussed with reference to English as an International Language, TEIL and WE (outlined in Hassall, 1996a & 1996b).

## INTRODUCTION<sup>2</sup>

Janicki (1985) calls for more research into the field of sociolinguistics particularly in the area in which the language learner deviates or is perceived by the native speaker to deviate from the norm. The present paper proposes that a consideration of relative deviance between different varieties of English may provide some indication of the relationship existing between such varieties. The term relative deviance has been used to refer to proximity to the norms of the standard language (cf. Romaine, 1982: 2). Kachru observes that generally 'the mother English' of Britain has been utilised as a norm against which to mark deviations although occasionally, as in the Philippines, Standard American English is treated as the norm (Kachru, 1990:135). In the present study it is considered that in order for a conceptualisation of English as an international language (EIL) to be developed, it is necessary to consider a more objective measure of relative deviance between varieties that is more of a two-way phenomenon. This international relative deviance (relDEV.EIL) would not necessarily compare a variety to a standard language but would operate between any two varieties of a language.

The approach assumed in this paper is one designed to support the pedagogy of

---

\* Faculty of Applied Studies, International Pacific College, Private Bag 11021, Palmerston North, New Zealand. phassall@ipc.ac.nz

\*\* IIST(Statistics), Massey University, Private Bag 11222, Palmerston North, New Zealand. s.ganesh@massey.ac.nz

TEIL outlined in 'Implementing EIL: the medium really is the message' (Hassall, 1996a) and 'Where do we go from here? TEIL: a methodology' (Hassall, 1996b). The methodology of TEIL encourages increasing the awareness of EIL and WE as a valuable resource and suggests that interactants utilising different varieties of English should attempt intercultural communication even though they may not be fully conversant with each other's codes (cf. Strevens, 1980). TEIL has been developed from Smith's basic inclusive principle: "English is the property of its users native and non-native, and all English speakers need training for effective international communication" (1987:xi). Hassall (1996b:421) discerns a distinction between prospective TEIL which is considered: "idealistic and innovative and may ultimately be concerned with the creation of new canons of English through negotiation between different users and varieties of world Englishes" and 'Teaching world Englishes' (cf. Kachru, 1989) which would appear: "largely retrospective describing how things are, from a multiple perspective" and which, if no alternative construct such as EIL is offered, may ultimately have to depend on the "identification of a standard variety of English for good communication between participants". World Englishes, as proposed by Kachru (1985) includes consideration of the Outer and Expanding Circles of English, in addition to the Inner Circle as represented by the Major Varieties of English (MAVEN) (Svartvik, 1997).

## **BACKGROUND TO THE STUDY OF DEVIANCE**

The concept of deviance is considered more idiosyncratic than variance (cf. Quirk, 1995), and is used when considering the language that results when, for instance, identical language tasks are undertaken by representatives of different language varieties. In order to determine relative variation between varieties more comprehensive corpora, representing the totality of the varieties would be required to be compared (cf. Mair, 1997). The present paper suggests something more modest and investigates deviance as a partial measure of variation. A conceptualisation of deviance is offered that integrates both negative deviance as arises in error analysis (cf. Corder, 1974) and positive deviance as occurs in literary studies and stylistics (cf. Leech, 1969; Van Peer, 1986). This unified notion of deviance conforms to Wales' (1989) view of deviance and

deviation which are, according to her, 'generally used synonymously, strictly referring to divergence in frequency from a norm, or the statistical average'. She claims that it is not surprising that 'statistical deviance' easily becomes associated with what is unusual, unpredictable, unexpected, unconventional. Crystal considers that there are different levels of deviance: "There are, moreover, different levels of deviance - degrees of departure from the norms which identify the various varieties of English ..." (Crystal, 1995:395). He also notes the significance of statistical deviance: "Slight degrees of deviance will hardly be noticed, or will produce an effect which it will be difficult to pin down. For example, an increased use of a certain kind of vocabulary may become apparent only after a great deal of statistical investigation (as in the case of authorship studies)" (Crystal, 1995:395). Together these two parameters presented by Crystal (1995) - that deviance may identify various varieties of English and that slight degrees of deviance only become apparent through statistical investigation, epitomise the approach that is to be elucidated in the present study. These are applied to establish the identity of corpora compiled from texts created by different groups of users of WE varieties.

Rather than direct contrast with standard Englishes (SBE or GA) which would result in a standard relative deviance (relDEV.SBE or relDEV.GA), contrast between the language user groups is made by comparing the language of each of the groups with the aggregate of the groups involved in the interaction, which may be considered to represent a 'notional EIL'. This characterises the international relative deviance analysis (relDEV.EIL).

### **ANALYSIS OF INTERNATIONAL RELATIVE DEVIANCE**

Hassall & Ganesh (1996) outlines an empirical study of relDEV.EIL that involves a comparative study of the language produced by different groups of English users. In that study an identical language task was given to all individuals in four groups of "English as an International Language (EIL) users" as identified by the analyst - an EIL practitioner. The language task involved a poem by Vernon Scannell entitled "Incendiary", about a fire at a farm (Godwin's Farm), which was to be rewritten as a newspaper article. The aim was to contrast between the four groups with reference to

the written language they produce rather than the ability they display in producing language as close as possible to a standard native English speaking norm.

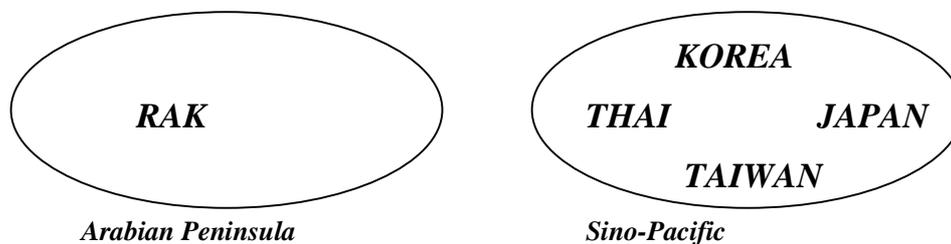
The present study undertakes to differentiate the language produced by English language users who are tertiary students in very different 'Outer Circle' countries (cf. Kachru, 1989: 15-16). It employs correspondence analysis to compare electronic corpora compiled from the written English of groups of tertiary students in different countries whilst completing a language task modelled on the written test of the International English Language Testing System (IELTS):

**WRITING TASK**  
You should spend no more than 40 minutes on this task.  
**TASK:** Write an academic essay entitled:  
**"The advantages of living in a large city."**  
You should write at least 250 words.

As in Holmes et al. (1991:23), a network technique was employed to locate appropriate key individuals who had sufficient contacts and expertise to arrange the sampling procedure. In order to differentiate the language produced by language users from different countries it was intended that the groups should be fairly homogeneous comprising tertiary students having similar backgrounds, nationality and sharing the same first language. Students were encouraged to complete this academic writing task to the best of their ability. An individual response was required without preparation or dictionary assistance. In 1997, one of the analysts, as EIL practitioner, was utilised as the primary node in the network and respondents were drawn together to comprise the **THAI, TAIWAN, RAK, KOREA** and **JAPAN** language user groups. **RAK** refers to Ras al Khaimah which is one of the seven emirates that comprise the United Arab Emirates (UAE); **KOREA** refers to South Korea. A further set of data was collected and mailed from Nigeria; however, this failed to reach New Zealand where the data was collated and analysed. There were 31, 45, 31, 100 and 94 individuals in the groups **THAI, TAIWAN, RAK, KOREA** and **JAPAN** respectively.

There is no immediately obvious relationship between the five groups of WE users involved in the study. Present and past contact with other countries and peoples who use, or have used, English varies considerably (see e.g. Kachru (ed.), 1992 ). Apart

from the physical proximity of Japan and Korea, the countries are widely dispersed geographically. The Arabic language users of **RAK** have their own characteristic writing system and are also likely to be acquainted with Farsi, Urdu and Hindi, as are others in neighbouring states. The **THAI**, **TAIWAN**, **KOREA** and **JAPAN** language user groups are all, to a greater or lesser extent, likely to be conversant with Chinese characters in addition to having their own distinctive orthography. The latter four groups could be characterised as ‘Sino-Pacific’ as distinct from the ‘Arabian Peninsula’ **RAK** group. Intuitively, one would expect some sort of division between the **RAK** group and the ‘Sino-Pacific’ groups. Apart from this however, it would be extremely difficult to predict any relationships between the groups even when considering so-called ‘ability’ or nearness to ‘native-speaker norms’. Hence:



In contrast to an 'emic' approach as proposed by Pike (1964), the statistical approach suggested by an international relative deviance analysis as outlined above, involved 'etic' principles as assumed in phonetic and graphetic analysis, where the physical patterns of language are described with a minimum of reference to their functions within the language system. No mediating parameters were applied and all responses to the language task were included, since it was considered that, together, these reflected the corpus that the EIL practitioner would have to deal with in the classroom.

The computer program "WORD" (Nation et al. 1988), which identifies words as being separated by spaces, full stops and apostrophes, was used to produce frequency listings of each word type for each of the four groups. These word types should perhaps more correctly be referred to as orthographic units, since the boundaries between the items are rigorously applied. The frequency list therefore not only includes words that are considered well-formed in standard English but also considers all orthographic

forms that are bounded by spaces or punctuation. These are familiar to both teachers and students of English since they are the focus of text manipulation computer programs including cloze and concordance software such as Fun With Texts (Davies, 1985) and Concord (Kennedy, 1991). These orthographic units comprise the written language that the teacher, and also the students, in an EIL classroom have to try to interpret and deal with. Frequency of orthographic units reflects the end product in terms of the language created by the totality of the students in each group in response to the task. Kenny (1982:66) refers to what we have categorised as an orthographic form as an 'unlemmatised word'. He claims that, in a sense, unlemmatised word counts contain more information since it is always possible, with some effort, to construct lemmatised word counts from unlemmatised word counts but the converse is not possible. When dealing with comparisons across language varieties, particularly when dealing with 'lesser known' varieties of English that are inadequately codified or described, lemmatisation is not yet possible and it is only feasible to consider unlemmatised orthographic units.

Study of the relative frequencies was undertaken by sorting all the data with respect to the aggregate group representing the totality of the data. Surface items that are not shared are obviously anomalous and deviant between varieties. Motivation for study of such exclusivity is useful for establishing the identity of individual WE. Equally important, when considering English as a medium for international communication, is study of the items that are shared by all of the varieties. Deviant usage relating to shared items is a delicate feature and for this an approach was made to the multidimensional statistical technique correspondence analysis as introduced by Greenacre (1984) and Ganesalingam & Lai (1994). CA facilitates dimensionality reduction and provides graphical displays in low-dimensional spaces. In other words, it converts the rows and columns of a data matrix (contingency or frequency table) into a series of points on a graph. For further direction in the practical utilisation of the technique and the background to CA in linguistics refer to Hassall & Ganesh (1996).

Comparison of each of the separate varieties with the totality of 'notional EIL' as represented by the aggregate EIL group (sum) enabled a two way contingency table to be assembled. In total there were 51,602 word counts distributed over 3,446 levels of

orthographic units and five levels of groups of text users. A mean value of 10,320.4 word counts per text group was calculated.

## **CORRESPONDENCE ANALYSIS AND INTERPRETATION**

This initial study examines only the first nineteen of the most frequent of the 3,446 orthographic unit/word types. Each of these appeared at least 500 times when all five groups were 'pooled' together. A contingency table of these words against the five groups is shown in Table 1. The analysis was carried out using the CORRESP (and GPLOTT) procedure(s) of the SAS System (1990).

**Table 1a.<sup>3</sup> Contingency table showing the nineteen most frequent word types distributed against language user groups.**

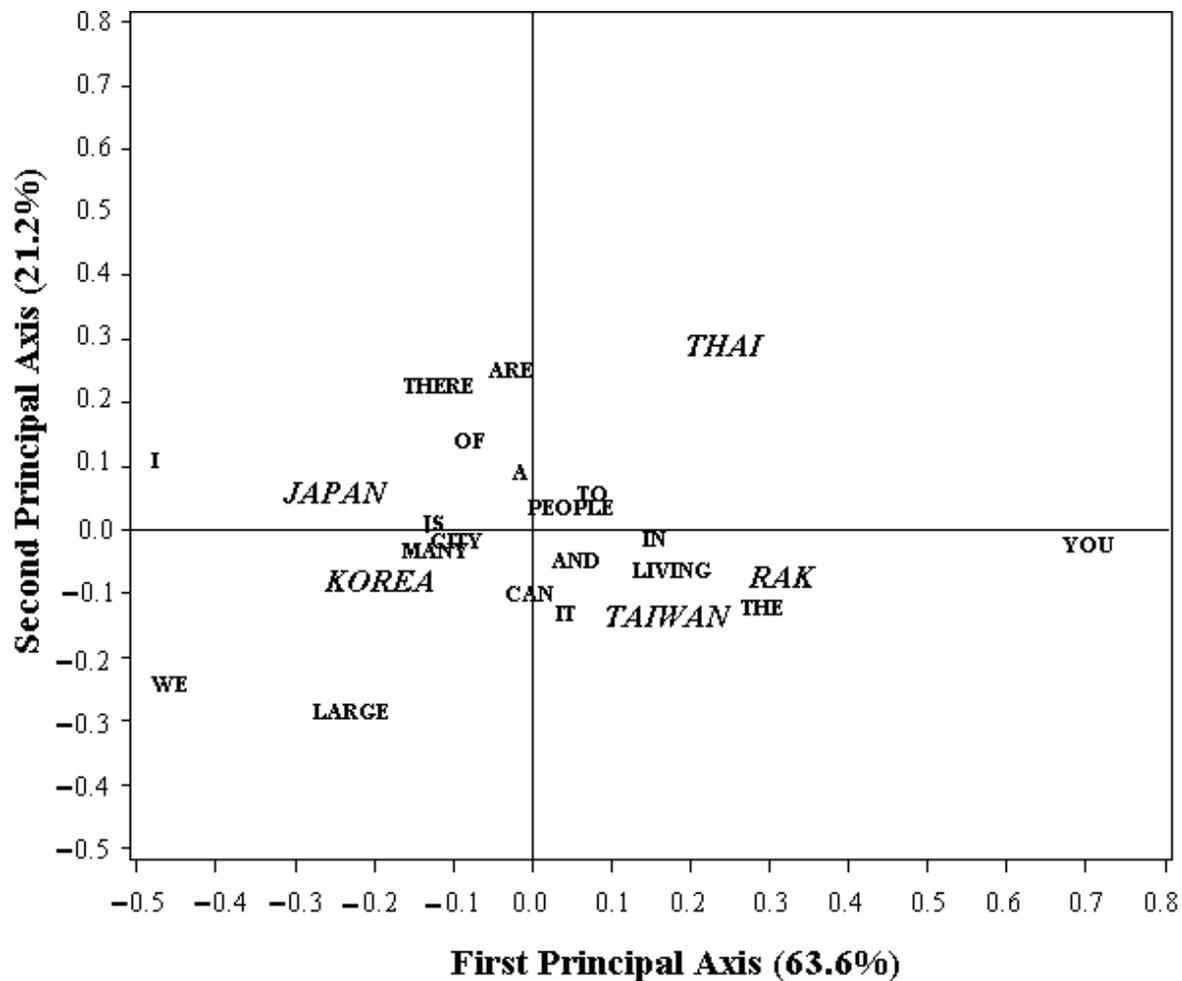
Word	<i>THAI</i>	<i>TAIWAN</i>	<i>RAK</i>	<i>KOREA</i>	<i>JAPAN</i>	Sum
IN	279	468	373	569	392	2081
A	279	389	237	559	493	1957
THE	228	528	359	461	260	1836
CITY	184	340	206	573	433	1736
AND	150	254	273	451	309	1437
TO	182	305	159	336	293	1275
LARGE	1	259	163	449	342	1214
OF	183	182	121	389	275	1150
CAN	94	210	181	346	233	1064
MANY	111	140	135	408	194	988
IS	101	187	105	310	260	963
ARE	175	120	119	282	230	926
I	62	110	42	276	377	867
WE	7	141	34	375	223	780
THERE	117	90	103	198	245	753
YOU	124	216	213	66	65	684
PEOPLE	90	118	98	209	127	642
IT	42	127	84	147	118	518
LIVING	71	140	69	151	75	506
Sum	2480	4324	3074	6555	4944	21377

A major aim of undertaking the correspondence analysis is to determine whether the five groups may be differentiated solely with reference to the relative frequencies of the different word types across the language groups. Once the contingency table is presented to CA, the procedure yields a conditional expectation for each row-column

combination of categories similar to that of “Chi-Square test of independence”. These values are normalized, and then a process much like Principal Component Analysis (PCA) defines the lower-dimensional solutions. The *total inertia* (similar to that of the *total variation* in PCA) is decomposed to represent the new dimensions. This total inertia is directly proportional to the Chi-Square statistic (for test of independence) and is a measure of total variation of the elements in the table. The number of maximum (new) dimensions obtainable equals  $\{\min(\text{no. of rows, no. of columns}) - 1\}$ . The low dimensions then simultaneously relate the rows and columns as points in a single plot. The axes of this low-dimensional configuration are called '*principal axes*' and are arranged so that the first principal axis accounts for most of the inertia, the second explains the second largest percentage of inertia and so on. It should be noted here that the plot should be thought of as two different overlaid plots, one for each categorical variable (i.e. rows and columns). Distances between category-points within a variable (i.e. distances between rows or between columns) have meaning, but distances between category-points from different variables (i.e. between a row and a column) do not. An important point to note in the plots is that the points that lie closer to the *origin* with respect to a principal axis contribute very little to the inertia explained by that axis. Thus, a principal axis can be ‘characterised’, even given a title, depending on which categories of rows and/or columns contribute the most to that axis.

The decomposition of the *total inertia* showed that the first principal axis accounts for about 64% of this total inertia followed by a 21% accounted for by the second principal axis. In other words, about 85% (= 64% + 21%) of the information can be accounted for by the first two principal axes, thus the association between the 19 word types considered and the five groups is mainly two-dimensional. Noting that the maximum number of new dimensions in this study is 4, i.e.  $\min(19,5) - 1$ , completeness of description provided by the extra dimension (accounting for just 15% of the total inertia) would be very much at the expense of clarity and ease of reference.

**Figure 1. Graphical display of the CA for the nineteen most frequent (>500) word types across the five language user groups**



The correspondence between the 19 words and the 5 groups is displayed graphically in Figure 1 for the first two principal axes. Note that, in this display two sets of points are *super-imposed*, one representing the word types and the other representing the language groups. The order of magnitude between the groups and amongst the word types is apparent in this display. The variation from left to right (along the first principal axis) opposes the *JAPAN* and *KOREA* groups, against the *RAK*, *THAI* and *TAIWAN* groups with the major contribution being undertaken by

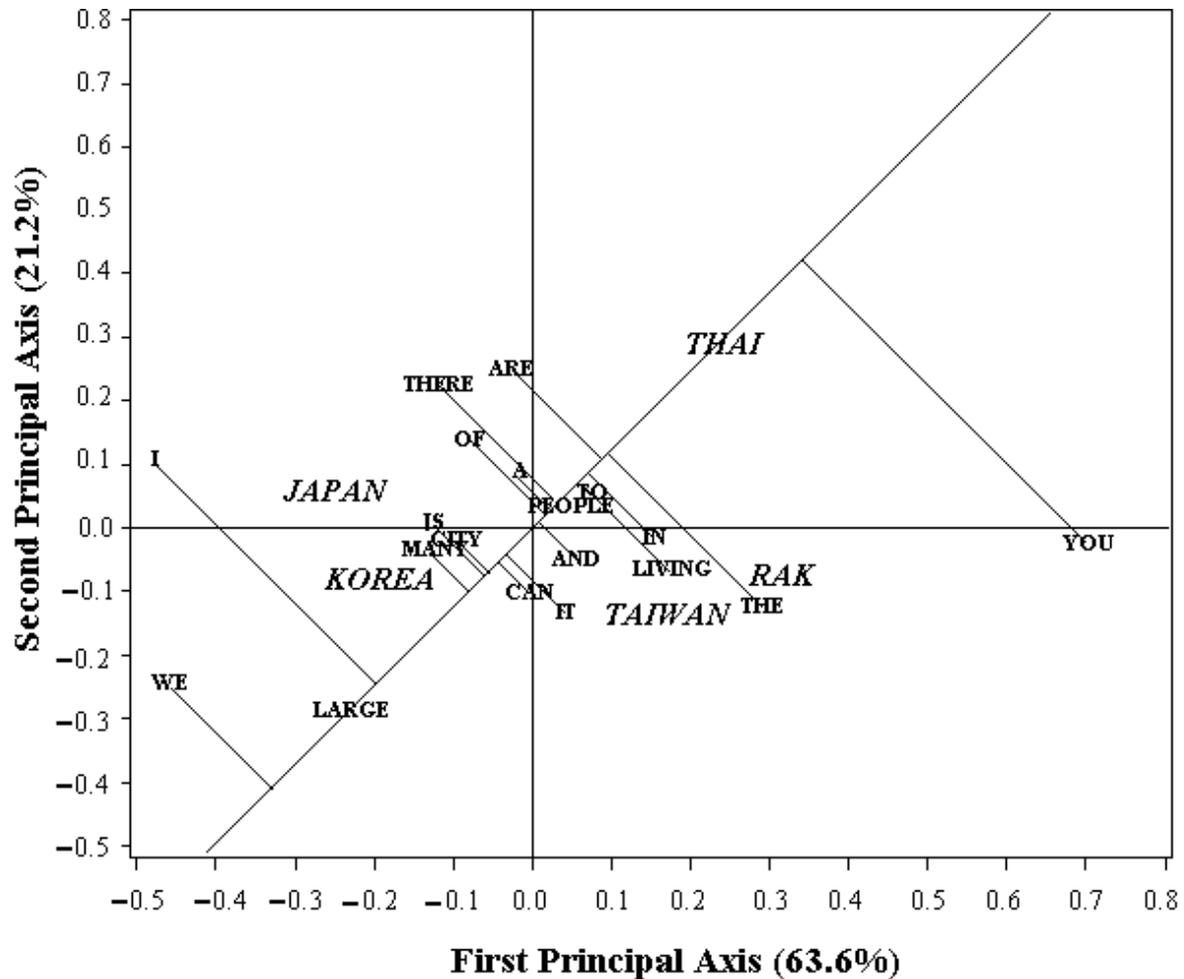
*JAPAN* and *RAK*; whereas the variation from top to bottom essentially opposes *THAI* and *TAIWAN*, with the other groups providing a lesser contribution. The CA provides a two-dimensional representation of the relationship between the four groups and suggests that the greatest contrast is,

***JAPAN/KOREA ↔ RAK/THAI/TAIWAN***

Basing an analysis upon the CA of the orthographic forms provides a more accurate description of the synchronic, surface relationship between the language user groups than the intuitive representation (shown on page 5 above) that hypothesised a likely contrast between the Arabian Peninsula and Sino-Pacific groups.

The behaviour of word-types reveals a contrast between words such as **WE** and **I**, and **YOU** along the first principal axis. The words **LARGE** and **THE** also make a moderate contribution to this contrast. The second dimension, however, accounts mainly for the differences between words **LARGE, WE** versus the words **ARE, THERE**. Although, the above behaviour of student-groups and word-types on their own may be useful, the interdependence of these two categories is also of interest in this study. The general concept is that, a particular column profile would tend to fall in a position which corresponds to the row categories which are prominent in that column profile. For example, the *RAK* point lies furthest on the positive side of the first principal axis and any word types that lie on the positive side of the first axis (i.e. words such as **YOU** and **THE**) could be regarded as ‘influential’ for this group.

**Figure 2. Projections of the nineteen most frequent (>500) word types onto the secondary axis through the *THAI* group**



The relationship between row and column profiles may be examined by considering projections of the ‘word-type’ points onto line(s) drawn through the ‘group’ point(s) and the origin on the graph. This in turn, enables us to relate, for example, the *THAI* group with all 19 word types as shown in Figure 2. Table 2 elaborates on this information and provides lists of the 19 most frequent word types in the aggregate corpus ordered so that those exerting the most positive influence

(attraction) on a language user group appear at the top of each list whereas those contributing a negative influence (repulsion) appear at the bottom.

**Table 2. Influence of the 19 most frequent word types on language user groups**

<i>THAI</i>	<i>TAIWAN</i>	<i>RAK</i>	<i>KOREA</i>	<i>JAPAN</i>
YOU	YOU	YOU	WE	I
THE	THE	THE	I	WE
ARE	LIVING	LIVING	LARGE	THERE
IN	IN	IN	MANY	LARGE
TO	IT	IT	IS	IS
LIVING	AND	AND	CITY	OF
PEOPLE	CAN	TO	CAN	MANY
THERE	TO	PEOPLE	THERE	ARE
A	PEOPLE	CAN	OF	CITY
OF	LARGE	A	IT	A
AND	CITY	CITY	A	CAN
IT	A	ARE	AND	PEOPLE
CAN	MANY	MANY	PEOPLE	TO
CITY	IS	OF	ARE	AND
IS	OF	IS	TO	IT
MANY	ARE	LARGE	IN	IN
I	WE	THERE	LIVING	LIVING
LARGE	THERE	WE	THE	THE
WE	I	I	YOU	YOU

An examination of the most frequent orthographic units used by each of the language user groups enables slight degrees of deviance to be observed. It can be seen from Table 2 that the *TAIWAN* and *RAK* groups are drawn towards use of the six words **YOU, THE, LIVING, IN, IT, AND** in that order. The *THAI* group (Figure 2) shares some similarities with three of the same words **YOU, THE, IN** appearing at identical rankings and a further word **LIVING** arising in the list but appearing slightly less influential. These words contribute negatively to the behaviour of the *JAPAN* group and to a lesser extent to the *KOREA* group.

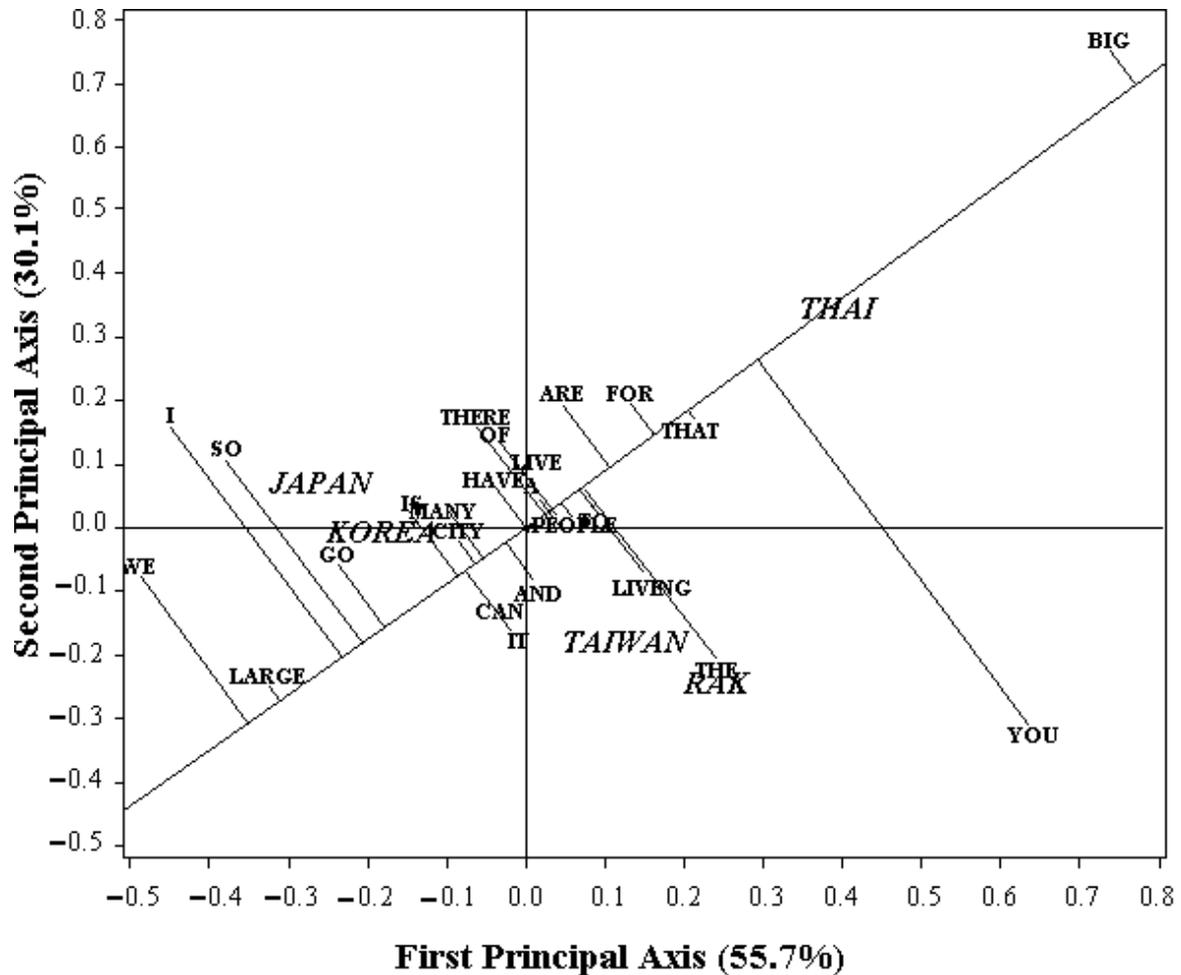
In contrast, the groups *KOREA* and *JAPAN* may be characterised by the positive influence of a different set of words. The *KOREA* and *JAPAN* groups share four of their six positively influential word types **WE, I, LARGE, IS** but in a different order and the other four words **THERE, OF, MANY, CITY** are not positively influential for the *TAIWAN, RAK* and *THAI* groups. The influence of these two major

groupings of words explain the primary contrast between **KOREA** and **JAPAN** as opposed to **THAI**, **TAIWAN** and **RAK**.

A secondary contrast can also be seen in Figure 2 (and Table 2) when examining the influence of the words **ARE** and **THERE**. These two words are listed in the eight most positively influential words for both the **THAI** and **JAPAN** groups but appear in the last eight (negatively influential) words for both the **TAIWAN** and **RAK** groups. Their appearance in the **KOREA** group is inconclusive with the item **THERE** appearing in the first eight (as **THAI** and **JAPAN**) and the item **ARE** appearing in the last eight (as **TAIWAN** and **RAK**).

In general it would appear that when considering the nineteen most frequent word types only, certain language groups are *attracted by* particular word types and *repelled by* others. Some words however contribute in a similar way to each of the language user groups, in particular the word types **A**, **CAN**, **PEOPLE**, **TO** are positioned similarly for all five groups and hence exert neither a positive nor negative influence. Productive inquiry into relative deviance might best be achieved by examining the outliers such as **I**, **WE**, **LARGE**, **THE** and **YOU**.

**Figure 3. Projections of the twenty six most frequent (>300) word types onto the secondary axis through the *THAI* group.**



Taking a wider view, Figure 3 provides a representation of CA for the 26 orthographic units that occur more than 300 times in the aggregate corpus. It would appear that, when considering the most frequent word types, there is attraction between certain word types and language user groups and repulsion between others. Examination of the word type **LARGE** warrants further scrutiny. Out of the 6,613 orthographic units used by the *THAI* group, **LARGE** occurs only once. Close scrutiny of the opening lines of the participants' scripts, reveals that the *THAI* group responded

to the stimulus “*a big city*” rather than “*a large city*” in the wording of the language task, provided by the ‘key individual in the network’ (their lecturer – page 4 above). This anomaly in the framing of the question is evident from an examination of Figure 3 which shows the orthographic unit **BIG** (here synonymous with **LARGE**) has been utilised primarily by the *THAI* group. Out of the 392 occurrences in the aggregate corpus, the word **BIG** has been used 190 times by *THAI*, one of the smallest of the language user groups. The *RAK* group, which has the same number of participants as the *THAI* group, displays a frequency of only 25 for **BIG**. In Figure 3, the positions of the units **LARGE** and **BIG** in relation to the language user group *THAI* demonstrate how the representation of correspondence analysis may best be interpreted. Generally, ‘overuse’ of a particular item by one language user group when compared to all the other groups will result in that item lying in a position further from the origin than the point referring to the language user group (as **BIG** in relation to *THAI*). Under-use of an item by one language user group, when compared to the aggregate of the other groups, will result in that item being positioned directly opposite the language user group, on the other side of the origin (as **LARGE** in relation to *THAI*).

Applying a similar interpretation to examination of the items **WE**, **I** and **YOU** provides more appreciable results for the study of relative deviance. From Figures 1 & 2, it may be seen that, when compared to other groups in the interaction, *JAPAN* and *KOREA* tend to under-use the item **YOU** and overuse the items **I** and **WE** compared to the other groups - with *JAPAN* tending to use **I** more than other groups and *KOREA* tending to use **WE** more than other groups. This may be contrasted with the *RAK*, *THAI* and *TAIWAN* groups which, when compared to the others, under-use both **WE** and **I** and tend to overuse **YOU**. This observation may be confirmed by examination of the data in Tables 1 & 2. This suggests a scale of objectivity and inclusiveness of the use of the generic pronoun in expository writing may be proposed, with **YOU** (favoured by *RAK*) being the most objective and not necessarily including the user, followed by **WE** (favoured by *KOREA*) being less objective but the most inclusive requiring inclusion of the user, followed by **I** (favoured by *JAPAN*) being the least objective and excluding all participants but the user (cf. Greenbaum, 1996:172; Leech & Svartvik, 1994: 58). This may be confirmed by an examination of concordances of each

orthographic unit for each language user group, see Figure 4.

**Figure 4. Concordances from the aggregate corpus indicating relative deviance**  
*Single citation indicates under-occurrence/ /Double citation indicates over-occurrence*

#### **Word = I**

**THAI:** stance, I can find everything that I want to buy in the department stores  
**TAIWAN:** t. When I need to go somewhere, I can take a taxi or bus. In a big city  
**RAK:** my time in towns or countryside. I am going to write about the advantages  
**KOREA:** e to live in a large city because I can be open to the opportunities, fir  
**KOREA:** lly one of the advantages is that I can learn proper Korean accent. In t  
**JAPAN:** ple. When I lived in my hometown, I couldn't buy anything after 8 p.m. Be  
**JAPAN:** ntages of living in a large city. I think one of the advantages of living

#### **Word = WE**

**THAI:** e public utility. For example, if we go to the provinces, we rarely find  
**TAIWAN:** ors in a big hospital. Finally, we can enjoy convenient life. I think t  
**RAK:** the best way for shopping, because we can find everything just in one buil  
**KOREA:** In a large city, until midnight, we can go home by subway or by bus safe  
**KOREA:** any libraries in a large city and we can read many books conveniently. T  
**JAPAN:** tion in a large city. Everywhere we go, we can see many advertisements of  
**JAPAN:** sn't cost much. In a large city, we don't have much parking space, so we

#### **Word = YOU**

**THAI:** s a center of transportation. If you are in London, you can choose many  
**THAI:** ngineering from a big university, you can start your salary at 20,000 baht  
**TAIWAN:** the 24-H convenient stores which you can find anywhere. If you want to  
**TAIWAN:** a convenient and colorful life, you can choose to live in a large city.  
**RAK:** tion when you live in a large city you can have a good job, so you will not  
**RAK:** e friends with different people so you can improve your language. Another  
**KOREA:** le. We have neighbours they help you when you are bothered. You help yo  
**JAPAN:** convenience store, and so on. So you can buy anything if you have money.

(utilising Concord developed by Kennedy, 1991)

### **CONCLUSIONS**

This study has examined deviant usage of shared items between language user groups representing different world Englishes, when making a comparison to the rest of the EIL community involved in identical interaction. In terms of the EIL user groups involved in this language activity, the high frequency of occurrence of **I** and **WE** in Japanese and Korean expository writing is clearly deviant when compared to the language of other groups involved in the interaction.. Similarly the frequent use of **YOU** by the Thai, Taiwan and Ras al Khaimah groups is plainly deviant, in this particular linguistic context, when compared with the language produced by the Japanese and Korean groups. This suggests a characteristic of each of the language user groups and points to a possible source of miscommunication. It will be seen that

orthographic units taking up the position of the word **BIG** in relation to *THAI* (i.e. along a secondary axis through a language user group and distant from the origin) is likely to provide the most anomalous situation where a group displays considerable deviance compared to the remainder of the EIL community, although here this may be regarded as contingent upon the vagaries of the empirical procedure (as explained on page 15).

While not fully representative of the World Englishes in question, this study has examined real data produced by groups of participants in different countries. As such it concurs with one of the guiding principles of Benzecri, who first developed the geometric form of CA within the context of linguistics, “The model must fit the data, not vice versa” (Benzecri, 1977). In order for this kind of analysis to be made applicable to *vernacular* World Englishes arising in different countries, it would be necessary to arrange the sampling procedure around key individuals who had access to a wider range of participants outside academia. For a more extensive study of the relationship between different ‘academic’ WE, the study could be repeated with data from the Writing Module of the IELTS or TOEFL tests. This would enable further reliable samples of expository academic English to be compared. If data from Inner Circle countries were examined as well as data from Outer and Expanding Circles, it would be possible to provide instances of deviant language usage even by so-called ‘native-speakers’ from the Major Varieties of English (Svartvik, 1997) when compared with the rest of the EIL community. This study has examined the linguistic variables specified as the most frequent orthographic units shared by the respondents. A similar methodology could be used to undertake investigations into features of spoken English.

No stereotypical explanations for the deviations, based on linguistic factors such as interference from the first language or non-linguistic cultural differences have been provided in this analysis. Rather, it is suggested that in the present naïve approach to the study of EIL, observation of these differences will be sufficient to instigate meaningful debate between the participants (and others) about reasons for such differences and point to fresh areas for investigation, such as the clear demarcation in the use of **I WE YOU** between the present WE groups. Hence the study of relative deviance, as shown here, may serve as an adjunct to research which relies more on the

extrapolation of prominence (Halliday, 1971:343) in a given variety, as identified by the observer/analyst, which may best be undertaken by those with direct access to the contextualised WE varieties in question.

## **IMPLICATIONS**

When English is used internationally between users of different varieties there is a marked disparity between the frequency of forms used by the interactants involved in communication. If little is known about participants' varieties of English then frequency of shared items will become more significant. It is hoped that the approach outlined here will stimulate debate and further inquiry into the interactants' world English varieties.

The present study has developed the concept of deviance and focussed on frequency of orthographic forms across different world English varieties. Other studies in the literature have examined the assortment of language forms that are produced in particular world Englishes with respect to their situation/cultural context. For many EIL practitioners (including teachers and students) such studies do not present a realistic view of their individual linguistic perspectives since they frequently have only a limited view of world Englishes other than their own. Synchronic, etic study of international relative deviance, as it is approached in the present paper, both strengthens the ontological status of individual varieties and points to distinctions between varieties that warrant further investigation. It is considered that the objective of investigations into world Englishes should be to inform and direct interactants and prompt further debate in order to enhance communication across varieties. Following Sinclair's (1980) demarcation of multiple source linguistics into retrospective and prospective patterning, the paradigm of world Englishes (Kachru, 1990) may be considered an instance of retrospective patterning that examines the way that language has been used up-to-now in various varieties of English. It is suggested that investigations into world Englishes (including MAVEN) and study of their relationships with one-another should input directly into the paradigm of English as an international language (EIL) which may be considered an instance of prospective patterning that considers how communication across varieties might be best accomplished. This is the realm of TEIL proposed by

Hassall (1996a & 1996b).

## NOTES

<sup>1</sup> For their invaluable advice and assistance in helping compile data for this research, we would like to thank the following:

Maurice McFarlane (Palmerston North, NZ); Harumi Tanaka (Nagoya, Japan); Lyn Doole (Miyagi, Japan); Phillip R. Morrow (Nagoya, Japan); Uhn-Kyung Choi (Seoul, Korea); Su-chiao Chen (Taipei, Taiwan); Passapong Sripicharn (Bangkok, Thailand); Ray La Bonte (Ras Al Khaimah, UAE); Victor A. Awonusi (Lagos, Nigeria); Toru Tadaki (Manchester, UK).

<sup>2</sup> An earlier version of this paper was presented at the 4<sup>th</sup> *International Conference on World Englishes: Language, Education and Power*, December 19-21 1997, which was organised for IAWE by the Department of English Language and Literature, National University of Singapore.

<sup>3</sup> **Table 1b. The seven most frequent unique items that are exclusive to a single language user group.**

Rank/3446	Word	THAI	TAIWAN	RAK	KOREA	JAPAN	Sum
86	SEOUL				103		103
133	NAGOYA					59	59
140	BANGKOK	53					53
278	DUBAI			20			20
567	KARAOKE					8	8
606	EXHIBITION				7		7
684	PALACE				6		6
696	SMOOTH			6			6
710	ALIVE				5		5

These items consist mainly of local capital cities. It would be necessary to go down to rank 827, frequency of 4 to find an orthographic unit exclusive to **TAIWAN**, the item - **CHOSE**.

## References

- Benzecri, Jean-Paul (1969) Statistical analysis as a tool to make patterns emerge from data. In *Methodologies of pattern recognition*. Edited by Satoshi Watanabe. New York: Academic Press. pp.35-74.
- Corder, S. Pit (1974) The significance of learners' errors. In *Error analysis: perspectives on second language acquisition*. Edited by Jack C. Richards,. London: Longman. pp.158-171.
- Crystal, David (1995) *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.
- Davies, Graham (1985) *Fun with Texts*. Maidenhead: Camssoft.
- Ganeslingham, Selvanayagam and Chin Diew Lai (1994) A statistical analysis of

- profiles and problems of recent Chinese immigrants. *The New Zealand Statistician*, **29**, 24-36.
- Greenacre, Michael J. (1984) *Theory and applications of correspondence analysis*. London: Academic Press.
- Greenacre, Michael J. (1993) *Correspondence analysis in practice*. London: Academic Press.
- Greenbaum, Sidney, (1996) *The Oxford English grammar*. Oxford University Press.
- Halliday, Michael A. K. (1971) 'Linguistic function and literary style'. In Chatman, S. (Ed.) *Literary Style: A Symposium*. Oxford: Oxford University Press. pp. 330-368.
- Hassall, Peter J. (1996a) Implementing EIL: the medium really is the message. *New Zealand Studies in Applied Linguistics*, **2**,57-77. (ERIC Document Reproduction Service No. ED 413 734).
- Hassall, Peter J. (1996b) Where do we go from here? TEIL: a methodology. *World Englishes*, **15**(3), 419-425.
- Hassall, Peter J. (1996c) Correspondence Analysis of Deviance in EIL. Paper presented at the Third Conference of the International Association of World Englishes, Hawaii: the East-West Center.
- Hassall, Peter J. (1998) Unity in Diversity: towards an integrated paradigm of English as an International Language and World Englishes including MAVEN. *The major varieties of English. Papers from MAVEN 97*. Edited by Hans Lindquist, Staffan Klintborg, Magnus Levin and Maria Estling. Växjö: Acta Wexionensia. pp. 291-301.
- Hassall, Peter J. and Siva Ganesh, (1996) Correspondence analysis of English as an international language. *The New Zealand Statistician*, **31**(1), 24-33.
- Holmes, Janet, Allan Bell and Mary Boyce (1991) Variation and change in New Zealand English: a social dialect Investigation. Project Report to the Social Sciences Committee of the Foundation for Research Science and Technology. Wellington: Victoria University of Wellington.
- Janicki, Karol (1985) *The foreigner's language: a sociolinguistic perspective*. Oxford: Pergamon.
- Kachru, Braj B. (1982) Meaning in deviation: toward understanding non-native English texts. In Kachru, Braj B. (ed.) (1992). *The other tongue*. 2nd edition. Urbana IL: University of Illinois Press.
- Kachru, Braj B. (1985) Standards, codification and socio-linguistic realism: the English

language in the outer circle. In *English in the world: teaching and learning the language and literatures*. Edited by Randolph Quirk & Henry G. Widdowson. Cambridge University Press. pp. 11-30.

- Kachru, Braj B. (1989). *Teaching world Englishes*. *Cross Currents*, 16(1), 15-21.
- Kachru, Braj B. (1990) *The alchemy of English*. University of Illinois: Illini Books
- Kachru, Braj B. (1990) *World Englishes and applied linguistics*. *World Englishes*, 9(1), 2-20.
- Kachru, Braj B. (ed.) (1992). *The other tongue*. 2nd edition. Urbana IL: University of Illinois Press.
- Kennedy, Graeme.D. (1991) *Concord: concordance program for studying words or phrases from the Brown or LOB Corpora*. Victoria University of Wellington.
- Kenny, Anthony (1982) *The computation of style*. Oxford: Pergamon.
- Leech, Geoffrey N. (1969) *A linguistic guide to English poetry*. London: Longman.
- Leech, Geoffrey N. and Jan Svartvik (1994) *A communicative grammar of English*. 2<sup>nd</sup>.edition. London: Longman.
- Mair, Christian (1997) *Variation studies based on machine-readable corpora: state-of-the-art and prospects for the future*. Presented at the *MAVEN 97 Conference*. Sweden: Växjö University College.
- Nation, Paul I. S., Alex Heatley, Alex Ivopol & Hwang Kyongho (1988) *FVORDS, WORD & ZERO Programs for PCs using MS-DOS*. New Zealand. Victoria University of Wellington.
- Pike, Kenneth. L. (1964) *Language in relation to a unified theory of structures of human behaviour*. The Hague: Mouton.
- Quirk, Randolph (1995) *Grammatical and lexical variance in English*. London: Longman.
- SAS Institute (1995) *SAS/STAT, SAS/GRAPH, Version 6 edition*. Cary, NC: SAS Institute.
- Sinclair, John M. (1980) *Some implications of discourse analysis for ESP methodology*, *Applied Linguistics*, 1(3), 253-261
- Smith, Larry E. (ed.) (1987) *Discourse across cultures*. Hemel Hempstead: Prentice Hall.

Strevens, Peter (1980) *Teaching English as an international language: from practice to principle*. Oxford: Pergamon.

Svartvik, Jan (1977) *Varieties of English: major and minor*. Presented at the *MAVEN 97 Conference*. Sweden: Växjö University College

Van Peer, Willie (1986) *Stylistics and psychology*. London: Croom Helm

Wales, Katie (1989) *A Dictionary of Stylistics*. London: Longman.