A COMPUTER-AIDED STUDY OF BECKETT'S AND TRANSLATIONS

This article is based on a comparative analysis of Samuel Beckett's plays and translations. It is directed at newcomers, like myself, to the field of computational linguistics in the hope that my experience when first using the computer may be of some value to them.

The study in question examines Beckett's play *Waiting for Godot* in both its French and English versions. After careful deliberation, it was decided that those traditional methods of stylistic analysis which rely on instinct would not allow for sufficient detail and accuracy. A more scientific approach was therefore adopted with the aid of the computer.

In this study, two main tasks were required of the computer. The first was to list all the words in the text and to indicate how frequently each occurs. This enabled us to recognise and evaluate individual items, and more particularly, to compare them with items in the target language. The frequencies, too, were revealing, especially where they differed in the respective languages. These word-lists could therefore isolate lexical units for close examination. The second function of the computer was to produce a concordance

25

which placed each word in its context, ordered alphabetically, providing another service and helping the stylistician to see, at a glance, how given items were used in the text.

These two tasks could have been undertaken by a number of different programs available to the linguist.. The program used here was COCOA[1] for several practical reasons. It was possible to obtain quick and efficient advice on how to use the system[2]. Another major consideration was that the program could be run on a local machine. Lastly, there was the fact that COCOA is a relatively easy program for the beginner to understand.

One of the main considerations when doing a computer-aided comparison of texts in different languages is that the concordances should be compatible. There is no difficulty if the concordance is a result of the same program, based on the same criteria. The difficulty arises when two separately compiled concordances are used. It then becomes necessary to adapt the information so that it is comparable in the two versions.

In the French concordance of *Godot*[3], which I received in its completed form, a more complex program was used. It not only listed words in context, with their number of occurrences, but it identified both the part of speech and the inflexions of a particular word[4]. Whilst appreciating the qualities of such a concordance, I found it impractical with the resources available, to attempt such a project. The result was that the English concordance was not able to sort and label words according to their part of speech.

26

This had to be done manually for comparative purposes.

Related to this feature of the French concordance was the fact
that the word-count came in two parts. In one, the basic forms
were given as they would appear in dictionary entries, together
with the frequency which included all variations. In the second
part of the word-count, each variation was given separately with
its corresponding number of occurrences. This was very useful for
examining verbal tenses and other grammatical inflexions.

The English word-count, again, did not make such valuable distinc-
tions. It simply listed each word with the number of occurrences.
When comparing figures, it was necessary to establish the basic form
and then search through the alphabet for its possible variants.

The disparity between the two concordances stresses the need in
stylistics for more compatible concordances which could be used
for comparative purposes not only between languages, but between
the works of an author in a single language. It also indicates how
important it is for concordances to be based on widely available
texts.

The benefits of using the computer are, nevertheless, numerous.
Above all, a concordance enables one to distinguish items more
clearly and to see them in context. In all three areas of the
linguistic analysis, the Vocabulary, the Grammar and the Style,
the concordance proved helpful.

27

In the Vocabulary section, for instance, it was found that certain
words were very common in one language and not in the other.
This was established initially by the frequency count. It was disco-
vered that high-frequency words presented the widest variety of
meaning, since through use, the meanings multiply.
These polysemic words were given special consideration.

In order to select items for consideration, the respective alphabeti-
cal word-lists were consulted. The English list had to be edited,
i.e. verbs and nouns reduced to their basic forms (infinitive and
singular) in order to tally with the French list. The new lists
were then compiled, one for each language, of all the words occu-
ring at least twenty times, since according to the correlation sug-
gested above, they were the most likely to have several meanings.
For ease of handling, these two lists were divided into two groups :
nouns and verbs, still ordered alphabetically. The two words in
each language were compared with their equivalents in the translation.
If an item appeared in one language and not in the other, or if
it had at least three versions in the translation, it was singled out
for comment. The figure twenty may seem somewhat arbitrary,
but it represents a reasonable degree of frequency. The lists
below show the final set of items chosen.

FRENCH LIST

(Frequency of words occuring at least 20 times)

Nouns	Freq	Translation	Freq	Verbs	Freq	Translation	Freq
Arbre	31	Tree	31	Aider	21		
				Arrêter	67		
Chapeau	53	Hat	54	Attendre	59	Wait	59
Chaussure	27[1]	Boots	33	*Avancer	30	Advance	20
*Chose	29	-		Bouger	21	*Move	30
Corde	41[2]	Rope etc.	34	*Chercher	23	-	
*Coup	26	-		Croire	28	-	
Fouet	24[3]	Whip	20	Demander	46	Ask	35
*Heure	22	-		Déposer	24	-	
*Main	51	Hand	26	*Dire	154	-	
Nuit	21	Night	20	Donner	44	Give	47
*Oeil	22	-		*Ecouter	20	-	
Panier	22	Basket	23	Essayer	30	*Try	37
*Pied	31	Foot	20	*Lever	54	-	
Silence	119	Silence	118	Mettre	57	-	
*Temps	136	-		*Parler	25	-	
*Tête	45	Head	34	Partir	20	-	
				*Passer	36	-	
				Penser	41	*Think	65
				Prendre	40	-	
				Rappeler	20	Remember	22
				Reculer	26	-	
				*Réfléchir	22	-	
				*Regarder	103	-	
				Relever	21	-	
				Remettre	20	-	
				Reprendre	37	-	
				Retouner	23	-	

Savoir	81	Know	78
*Tendre	21	-	
*Tirer	33	Pull	22
*Tomber	41	Fall	32
Venir	65	*Come	82
.*Voir	106	See	63
*Vouloir	85	Want	36

NOTES

Translations are given where possible. Auxiliaries are omitted, because there is no single translation or the frequency of the translation is below 20.

- : does not appear in opposite list
* : considered under polysemy.

1. Difference in number is caused by fact that "to take off boots" = se déchausser, used several times in English.

2. Difference in number caused by fact that *corde* is translated by *cord* when it refers to trouser *cord*.

3, Difference in number caused by additional stage-directions in French text.

30

ENGLISH LIST

(Frequency of words occuring at least 20 times)

Nouns	Freq	Translation	Freq	Verbs	Freq	Translation	Freq
				Advance	20	*Avancer	30
Basket	23	Panier	22	Ask	25	*Demander	46
*Day	25	-		*Begin	23	-	
Foot	20	*Pied	31	*Come	82	Venir	65
Hand	26	*Main	51	Fall	32	*Tomber	41
Hat	54	Chapeau	53	Give	47	Donner	44
Head	34	*Tête	45	Get	57	-	
Night	20	Nuit	21	Halt	26	-	
*Pause	84	-		Hear	21	-	
*Place	33	-		Help	37	-	
*Right	26	-		Know	81	Savoir	78
Rope	34	Corde	41	*Look	89	-	
Silence	118	Silence	119	Make	31	-	
*Time	39	-		*Move	30	-	
				Pull	22	*Tirer	33
				Put	66	Mettre	57
				Remember	22	Rappeler	20
				See	63	*Voir	108
				Say	72	-	
				*Stop	54	-	
				Take	79	Prendre(Re-)	77
				Tell	55	-	
				*Think	65	Penser	41
				*Try	37	Essayer	30
				Turn	29	-	
				Wait	59	Attendre	59
				Want	36	*Vouloir	85

31

Each of the significant items was considered. Thus, for instance, in the French list, the first item to which attention was drawn was *chose* with features frequently in both versions, but more so in French. *Chose* with its equivalent *Things* is well known as a "rescue" word in conversation when the right expression fails to come to mind.

In *Godot, chose* is left to its normal vague role, and the English attempts to clarify the meaning, as in :

> Pour en faire tomber quelque chose :
> as though to dislodge a foreign body

in which the English conveys more physically the effort involved. This *chose,* so firmly wedged that it must be dislodged, carries more weight in "foreign body". We know it is a living thing which adds to our amusement. Similarly, Pozzo's rare and condescending piece of flattery :

> Qui sait, vous m'aurez peut-être apporté quelque chose

becomes even more turgid in the English :

> Even you, who knows, will have added to my store.

What is one blessing amongst so many ? "Store" is stripped of the humanitarian element which accompanies an act of generosity.

32

Pozzo's intention to sell his slave "pour en tirer quelque chose" - to get a good price for him - is his crowning act of selfish materialism, and this is borne out in no uncertain terms by the translation.

The more explicit rendering exposes the material aspect of the French *chose* without disturbing the fluidity of the discourse. The conversational tone of "Tu as raté des choses formidables" is admirably preserved in "You missed a treat". This final example serves to illustrate two of the dominant features of the French which are its greater tendency towards the colloquial (raté = missed) and towards the abstract. *Chose* on its own is vague and needs the support of the rest of the sentence, particularly the qualifying adjective, to acquire its full meaning.

A similar type of commentary was conducted on all the significant items, marked by an asterisk on the lists, to illustrate the differences between the two versions. These differences were due in part to the structure of each language, and in part to Beckett's use of the latter.

As far as the grammatical section is concerned, one of the services offered by the computer was a reverse order frequency list so that the verb endings could easily be detected[5]. Both versions were found to be predominantly in the present tense. Given the medium, which is an imitation of dialogue, this fact is hardly surprising. The characters, trapped in an eternal present are not allowed to cast their

33

thoughts into the future. Their only refuge is in memory, and again, Beckett refuses to yield to the past as a source of solace. This pattern applies to both versions of the play. In the case of the present tense, there is little discrepancy between the two languages, except where French makes use of what might be called an "atemporal present" where the time is left deliberately vague. In English, this is usually translated by a future tense as in :

    je reviens = I'll be back
    on a le temps = Time will tell
    tu les gardes ? = you'll keep them ?

but the effect is the same, namely that neither refer to any precise moment in time. The characters are engulfed in an eternal present. The situation is very close to that of *Huis Clos* where the characters are condemned to live together for ever. Larthomas describes it vividly as[6] :

    ... la durée que rien ne viendra interrompre... Il ne
    reste plus qu'à continuer. Le tragique nait tout à
    coup de cette terrible simplicité.

In the section on style, the concordance was particularly useful in helping to detect differences in the images and idioms. It was noticed, for instance, that French, in rendering physical expressions, refers to parts of the body, where English suggests these by the verb or by another means :

34

les yeux vagues = staring sightlessly before him
fait quelques pas affolés = casts about wildly.

French in this rare case, is more attentive to detail, referring directly
to the physical feature involved.

More often, Beckett is more specific in the translation

in accordance with the British desire for explicitness :

Présente les faits = speaks of a thief being saved
Ça creuse = stimulates the jaded appetite.

These are but a few examples of how useful the concordance was
in detecting divergences in the two versions. The type of listing
found in the concordance also helped to promote a methodical ap-
proach to the comparative study and to determine the themes -
such as waiting, insecurity, lack of communication - which run
through Beckett's work.

I think it would be fair to conclude that the information from the
computer made the task of comparison easier in the long run,
especially as the printed texts did not match and could not easily
be contrasted. It cannot be denied, moreover, that the accuracy of
the word-count strengthened the observations made by the stylistician.

Considering the time and effort involved in producing a concordance

35

even for a single short play such as *Godot*, I would add, however, that the use of the computer is better justified if extended to a wider project, eg. several plays by Beckett. The results obtained from *Godot*, I would suggest, are sufficiently encouraging to make a more ambitious project worthwhile.

Anna HARGREAVES

36

NOTES

(1) For details of COCOA and the procedure used for this project, see Appendix I.

(2) Dr. Schonfelder, Dept. of Computer Science, University of Birmingham, was chiefly responsible for introducing me to COCOA and running the program.

(3) The French concordance was produced by Dr. Paul Fortier, University of Manitoba, Canada.

(4) See Appendix 2.

(5) See Appendix 3.

(6) Pierre Larthomas, *Le Langage dramatique* (Paris : Armand Colin, 1972, p. 158.

37

APPENDIX I

Procedure followed in using COCOA

The first step was the preparation of the text which was fed into the computer. Certain features of the original text had to be taken into account :

1) that it is written as a sequence of characters in upper and lower case;
2) that italics represent the stage directions;
3) that words are normally separated from each other by spaces or punctuation marks, but in some cases, words are combined by hyphens and apostrophes.

All this information had to be coded into machine-readable form using only a restricted character set of 63 characters, with further restrictions that certain characters were reserved for specific functions e.g. $<$ $>$ to enclose text references.

+ / to denote continuation of text lines from one card to another, and to denote termination of a text line.

Concerning the first two items listed above, uppercase letters were coded by using a preceding ↑ , and stage directions were enclosed in single brackets ( ). These details were not, in fact, used in this

38

analysis. But in view of the time and effort involved in punching a text, it was important to include as much information as possible for the benefit of other linguists in studies in which this additional information is of importance. Thirdly, there is the question of punctuation, which appears as normal except in the case of apostrophies and hyphens. Unless otherwise specified, COCOA does not take these into consideration. For the purposes of this study, it was therefore necessary to separate the negative particle *n't* from the verb, and this also applied to the abbreviated inflections of auxiliaries such as *'d, 'll,, 's, 've.* As for hyphens, if they were used as a·punctuation mark, beginning and ending a phrase in apposition, they were preceded and followed by a space to separate them from neighbouring words. In all other cases, the hyphen was left to join two elements of a single word in the concordance, as in *half-hunter* or *will-power* because they form a single unit of meaning.

Next, a word must be said about text references which are used to provide selection information and location reference points in the text. They must be distinguished from the normal running words of the text. This is accomplished in COCOA by use of the reserved character angle brackets to enclose the characters that make up the reference. The general form of a text reference is an follows :

$$< \text{ key character } \Delta \text{ reference string } >$$

where the *key character* is a single alphabetic character which

39

denotes the type of reference, and the *reference string* is a string
of alphanumeric characters which distinguishes the particular refe-
rence. For example in this text, the reference keys used were T,
A, C, P, L. (1) These refer to Title, Author, Character, Page and
Line references. The < T Δ GODOT > and < A Δ BECKETT >
references occur only once at the start of the text and serve as
overall identification references. The < C Δ *initial* > references
precede the lines of text corresponding to the speech of each cha-
racter. The characters are distinguished by their initials :

< C Δ E >	- Estragon
< C Δ V >	- Vladimir
< C Δ P >	- Pozzo
< C Δ L >	- Lucky
< C Δ B >	- Boy

Once the punching was complete, the cards were fed into the ma-
chine, read and stored in the computer filestore. A command was
issued for a print-out of the text in which typing errors were de-
tected. It is essential to correct these errors before proceeding any
further. This was done through the GEORGE 3 editing system avai-
lable on the computer. There are a series of editing instructions
which enable the user to replace one string of characters by a new
one, or to insert a string which was accidentally omitted during the
punching. These instructions constitute an editing language which
must also be mastered. The editing instructions of GEORGE 3
allow the user to make detailed alterations to his file. The original
file was then deleted and replaced by a new, correct version of the

40

text. This is a tedious task, but a necessary one in the interests of accuracy. If the text is not word-perfect, the errors will affect the concordance.

The corrected text was then ready to be processed, and the next step was to supply the COCOA program with a small set of punched cards specifying the operations for the machine to perform. To run COCOA, thirteen cards are needed, each card representing a particular instruction. The various options are outlined in the COCOA Manual.

41

APPENDIX 2

KEYWORD	INFLECTION CODE	S C	FREQ
A$^+$		001	381
A$^-$ $^+$		002	4
AU	CONTR: A$^+$ET LE	002	104
AUX	CONTR: A$^+$ET LES	002	6
ABANDONNER	INF.	001	1
ABANDONNE/9	P.P. ADJ. M. SING.	029	1
ABANDONNES/9	P.P. ADJ. M. PL.	033	2
ABANDONNEES/9	P.P. ADJ. F. PL.	035	1
ABANDONNE	IMPER. SING.	097	1
ABDULLAH#	M. SING.	001	1
ABEL#	M. SING.	001	2
ABONDANTE	P.PRES. F. SING.	007	1
ABORD		001	11
ABRITER	INF.	001	1
ABSENCE	SING.	001	1
ABSOLUMENT		001	1
ABSORBE	PRES. IND. 3 SING.	041	1
ACACACACADEMIE #11	SING.	002	1
ACCABLEMENT	SING.	001	1
ACCABLE/7	P.P. ADJ. M. SING.	029	1
ACCENT	SING.	001	1
ACCORD	SING.	001	2
ACCOUCHENT	PRES. IND. 3 SING.	041	1
ACCRUES	P.P. ADJ. F. PL.	035	1
ACHARNE/7	P.P. COMP. M. SING.	013	1
ACHARNE	PRES. IND. 3 SING.	041	2
ACHEVE+4	IMPER. SING.	097	2

42

APPENDIX 3

198	2	ATTENTIVE	16
199	232	WE	20
200	1	GAZE	1
201	1	SCRUTINIZE	1
202	1	LEAP	5
203	2	RELIEF	1
204	1	STIEF	45
205	52	IF	1
206	1	SELF	17
207	1	ITSELF	11
208	274	OF	18
209	7	BEG	11
210	4	BIG	1
211	6	HANG	1
212	1	RUBBING	1
213	2	ADVANCING	1
214	1	BLEEDING	2
215	1	FOLDING	1
216	1	FINDING	1
217	1	ALLUDING	10
218	1	CHAFING	1
219	3	SAGGING	1
220	2	RINGING	7
221	1	SINGING	1
222	2	CLUTCHING	1
223	1	BRANDISHING	1
224	3	BREATHING	13
225	1	EVERTHING	20
226	1	SHAKING	1
227	2	KICKING	3

43