

# I Like Both Myself and Me

Martin Edwardes

University of East London

The reflexive nature of pronouns varies with person, but only in the first person does the subject always co-identify with the object. Thus *I like me* has effectively the same meaning as *I like myself*. Grammatically, *I like myself* is considered marked, while *I like me* is unmarked. However, is the unmarked form the most used?

This exercise started as a search for occurrences of *I like myself* and *I like me* on the Internet, to identify simple frequencies of use. It emerged as an issue from the back-to-back LAGB and BAAL conferences at Reading during September 2001, where the issue of reflexivity was dealt with directly or indirectly by several presenters. In the “Pragmatics and Anaphora” workshop at the LAGB conference, Yan Huang presented a paper on *Anaphora, Syntax and neo-Gricean pragmatics*, and Anne Zribi-Hertz presented on *Reflexivity and Disjoint Reference: Where Does Syntax Begin?* Both of these papers concentrated on the issue of reflexivity in language use, and both used the syntactic definition of marked and unmarked forms: in the case of English, a marked reflexive has *-self* added to the object. Also, during his talk at the BAAL conference Eddie Williams mentioned, in passing, that a business reflexive seemed to have arisen as an emphatic (e.g. *you can write to ourselves at the above address...*). He saw this as a phatic attempt to add authority.

A question was posed at the workshop as to whether the grammatically marked form was actually the most frequently used form and, if so, was it really the unmarked form? It was a question that appeared to be easily answerable by reference to corpora, and the largest corpus of all, the Internet, seemed ideal to test the question.

Using the Internet as a corpus has both advantages and disadvantages. It has the two great advantages of being enormous and of being unselfconscious – there appears to be little or no self-censorship going on, so syntactic forms encountered can be viewed as grammatical at least at the point of origin (the *I-language* level). However, any study using the Internet has to be aware of certain problems.

The first is that it is homogenous: first language and other language usage are difficult to separately identify, and there is no way of knowing the age of the writer or the time when the material was written. It can therefore only be used for the grosser type of study, where simple headcounts are significant.

The second problem is identifying the most suitable search engine. When this project began, in 2001, Alta Vista was the only viable choice. It provided a search of whole pages instead of just a header or keyword search; it allowed for phrase searches; it recognised spaces; and it was regularly updated. For precisely these same reasons the best choices would now be Google or AOL, Alta Vista has somewhat fallen behind in the race. However, the first studies were made using Alta Vista, so all subsequent studies also used it. This was done to ensure, as far as possible, comparative figures.

The search itself is problematic: there are limitations in the Internet search engines that have to be taken into account. First, most engines do not recognise punctuation or capital letters, so a search for *I like myself* would also find “these are the things my husband and I like. Myself, I also like...”. Second, the string to be found must be enclosed in spaces and double quotes. Without the double quotes the selection will include any site with any one of the words. Without the spaces the search will pick up partial strings, so that a search for *I like me* would also find “I like meat” or even “Ferengi like meat”.

The final problem is that the internet is constantly changing, so some of the sites listed no longer exist or have been altered. What is being sampled is, therefore a smeared image of what is actually out there, and not a snapshot.

It is still possible to get useful data out of the exercise if the limitations are taken into account. Two facts have to be acknowledged: first, the absolute figures produced for frequency utterance cannot be considered individually mathematically accurate; and second, some anomalies are bound to get through the checking process. To resolve this problem each case can be checked in detail, but this is impractical with large hit counts. Fortunately, with large hit counts any anomalies are evenly distributed through the data, so the figures are accurate enough for comparative analysis. In this exercise the figures mostly represent high-volume data, so statistical anomalies will tend to be smoothed out. However, it does mean that any volumes under about 100 cannot be considered as statistically relevant. In this exercise several totals under 100 were found: they are included for completeness but cannot be considered useful data.

It was originally intended to look at the full range of reflexive pronouns, but this proved impossible because the reflexive nature of pronouns varies with person. In the third person there is the probability that the subject is not co-identified with the object; so in *he likes him* the subject *he* is unlikely to be the same person as the object *him*. Reflexion has to be marked by a syntactic flag, the addition of *-self*. In the second person, it is possible to either separately identify individuals in a receiver-group, or to redefine the receiver in the middle of an utterance. Thus *you like you* can, through deixis, have a different meaning to *you like yourself*.

Only in the first person does the subject always co-identify with the object: *I like me* has effectively the same meaning as *I like myself*. In the plural there is the slight possibility that *we like us* could refer to two different groups containing the sender, but in the examples checked no cases were found where this meaning could be imputed.

The first data check in September 2001 proved that *I like me* is more common than *I like myself*. However, it was decided to extend the study to look at the past tense, the plural and the comparable antithetical statement *I hate me / myself*. This immediately uncovered an anomaly: in none of the cases, except *I like me*, was the grammatically unmarked form in the majority. The most extreme cases (other person, other tense) were, unfortunately, statistically insignificant – even the Internet isn't a big enough corpus – but the other figures seemed to show that the further you get from *I like me / myself*, the more common was the grammatically marked form.

In February 2002 a second check was undertaken, this time extending the study to include related forms of *love*, *dislike*, *don't like* and *do not like*. This gave a theoretical continuum of meaning – although, it must be stressed, this continuum is not mathematically significant. This time the tense effect was not tested. The second check showed the same effect as the first, and gave the additional piece of data that semantic “distance” from *I like me / myself* seemed to affect relative frequencies.

In December 2002 a third check was performed, combining the features of the first and second checks. The variations were: singular and plural, present tense and past, and the same continuum of meaning as used in the second check. This gave twenty four binomial sets of data (figure 1). Once again, the same effects were found as in the first two checks: changing verb, tense or number reduced the frequency of the unmarked form, and changing verb and tense, or number and tense, had a cumulative effect. Once again, changing verb, tense and number gave a statistically insignificant sample.

The figures for the third check have been expressed in graph form to make them easier to review (figures 2 and 3). The graphs show four series of data, where the numbers on the x axis correspond to the row numbers in figure 1. So the series of data are 2 to 7, 9 to 14, 16 to 21 and 23 to 28. On the first graph (figure 2) the data appears to show two of the four sets as almost-perfect series of Poisson distribution. This is an artefact of the data: the points in the

series are arbitrary and do not represent a mathematical relationship. If there is any scalar distance between the  $x$  axis points, then the distance between *I love me* and *I like me* is not comparable to the distance between *I do not like me* and *I don't like me*. Because of this the data should be treated mathematically as twenty four binomial distributions between marked and unmarked forms. Mathematically, the binomial distributions are not comparable; but linguistically, they are. This gives a problem for analysing the data: if we insist on a mathematically rigorous analysis then the results we can produce are banal. If, however, we attempt a more free-form analysis then our conclusions will have to be more tentative. The second option was selected as likely to be the most productive.

The first thing that had to be done was to try to normalise the data. To do this, the first series (singular present tense) was used as the base. This is somewhat arbitrary, but it is the most numerous set and it contains the constructs that started this whole exercise. It thus provides a reasonable baseline. The second task was to predict the outcomes of the other sets of data based on the first, baseline set. The figures in the figure 1 *predicted* column show the distribution that would have occurred if the only variable was the continuum of meaning, while figure 3 shows the variation between the *predicted* column and the actual (*unmarked %*) column. In this graph we can see that the first form, *love*, and the total of the third through fifth forms (*do not / don't / dis-like*) have notable negative values compared to the positive value of the second, base form, *like*. There do appear to be separate effects from verb, tense and number which cumulatively and negatively affect the frequency of the unmarked form. It can also be seen that the data in the fourth series (other tense / other person) is highly variable compared to the first three series. This is caused by the low total values in this series, and serves to emphasise the unreliability of this data.

What could be going on here? It seems unlikely that anything in the formalist canon could explain this, so solutions have not been sought there. In the functionalist area it could be a product of the textual metafunction, or possibly something to do with the logical metafunction. It could be that the frequency variation in constructs could be a feature of the constructs themselves, or it could be to do with the contexts in which they are found.

The logical metafunction is concerned with context and the relationship between utterances rather than the utterances themselves. It is associated with such terms as holism and idiom, inasmuch as they all three express, at some level, the use of a construct as a standalone agrammatical form. Although there may be apparent grammar operating within the construct, the production can be considered as unitary – what Alison Wray calls in another context, “performance without competence.” This was also the subject of a survey carried out as part of my MA, which showed that the same construct can be viewed in different ways by different individuals, so that what for some is an idiomatic unit is for others a grammatical composite. What is more, there seemed to be an individual tendency to view constructs in an idiomatic or grammatical way: some people adopted a grammatical approach when it seemed inappropriate, and vice versa. The idiomatic nature of a construct does not appear to be a grammatical or even cultural absolute.

If *I like me* is being used as an idiom, then what we may be seeing is a “contamination” from the idiomatic form *I like me* into related forms. As the distance from the original construct increases, so the idiomatic contamination reduces: the willingness to use the idiomatic form reduces as the influence of the idiom wanes. This explanation fits with the facts, and may well be a partial solution. However; it is unsatisfying as a full solution in that it does not include any context-related data.

To address this, the first 100 occurrences of *I like me* and *I like myself* were investigated to see what their context revealed. Of the 100 *I like me* occurrences, 79 were stand-alone constructs related to self-help products; 12 occurrences were standalone comments unrelated to self-help; 7 were embedded in larger constructs (e.g. *I like me best with a beard*); one was a dialect replacement for *my* (*I like me sex on the sharp side*); and one was elliptical (*I like me some Chris Moore*).

Of the 100 *I like myself* occurrences, only 6 were related to self-help products, and 64 were standalone comments unrelated to self-help; 26 were embedded in larger constructs (e.g. *I like myself at this weight*) and 4 were emphatics (e.g. *it's what I like myself*). Surprisingly, 45 were related to the personal weight of the writer, compared to zero for *I like me*.

Thus there seem to be pragmatic issues involved alongside the idiomatic issue, at least around the *I like me / myself* dichotomy. Could there actually be a subtle difference in meaning between the two forms? Anthony Giddens points towards issues of self which are related to the intimate self and a reified external self. I can see *me* as a single entity, with the reflexion back to that single entity, or I can see *myself* as two entities, with the reflexion moving from one version of *me* to the other. It gets more complex, because the internal *me* is related to the intellectual *me*, the *me* that drives the machine; while the external *myself* is related to the physical *myself*, the machine that is driven. There is more than a small hint of Cartesian dualism here. However, the figures that come out of this analysis do seem to relate to the *me* that can be liked (the intellectual) and the *myself* that can be weighed (the physical). Although Lakoff and Johnson don't mention this dichotomy of self directly, these could be two of the most fundamental metaphors we live by.

If this argument has merit then it should be possible to prove it by looking at the other constructs in the data collected. When this is done the conclusions do seem to apply to most of the other cases, although not all of them.

Figure 4 shows the context-related data for *I hate me / myself* and *we like us / ourselves*. Almost all of the *I hate me* constructs seem to be blanket condemnations of the self, while less than half of the *I hate myself* constructs are so severe. Sixteen of them directly mitigate the hatred by making it circumstantial. With regard to *we like us / ourselves*, all but one of the *us* constructs refer to a corporate *us* (*us* as an indivisible single group as opposed to *us* as a divisible set of individuals), and the liking is unconditional. The data for *we like ourselves* is more complex. All of the liking appears to be unconditional, but there is a difference between a corporate *ourselves* and the identification of *ourselves* as a group of individuals. In some cases in the source data the individuality was clear, but in others it was less so, so there may have been some misinterpretation of the data. However, there was nonetheless a majority of corporate *ourselves* rather than the predicted majority for individuals.

So what can be concluded from this? The main outcome seems to be that what looked a simple question at first sight has turned out to be a complex mixture of issues. What should have been a straightforward grammatical exercise has become mixed with idiom and the semantics of self to make a complex interpretive problem. It is an illustration of the difficulties that surround corpus examinations: to produce list of constructional frequencies is banal, but in analysing the frequencies the data fights back against being pinned down.

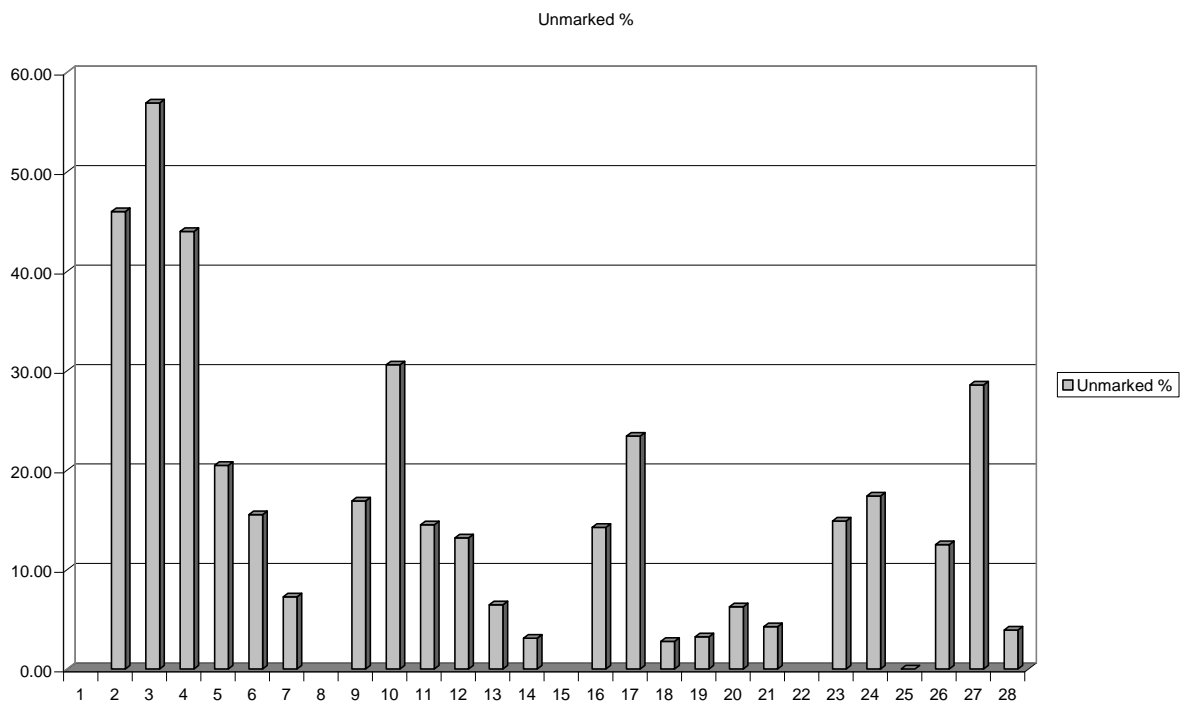
If the data is being read correctly then the conclusion would seem to be that a large part of the use of the unmarked construct *I like me* is a product of idiom. However, as the related forms become more distant from this exemplar construct, the more contextual effects come into play. *I like me* is used because it is idiomatic; *we hate us* is used because it has a different meaning to *we hate ourselves*.

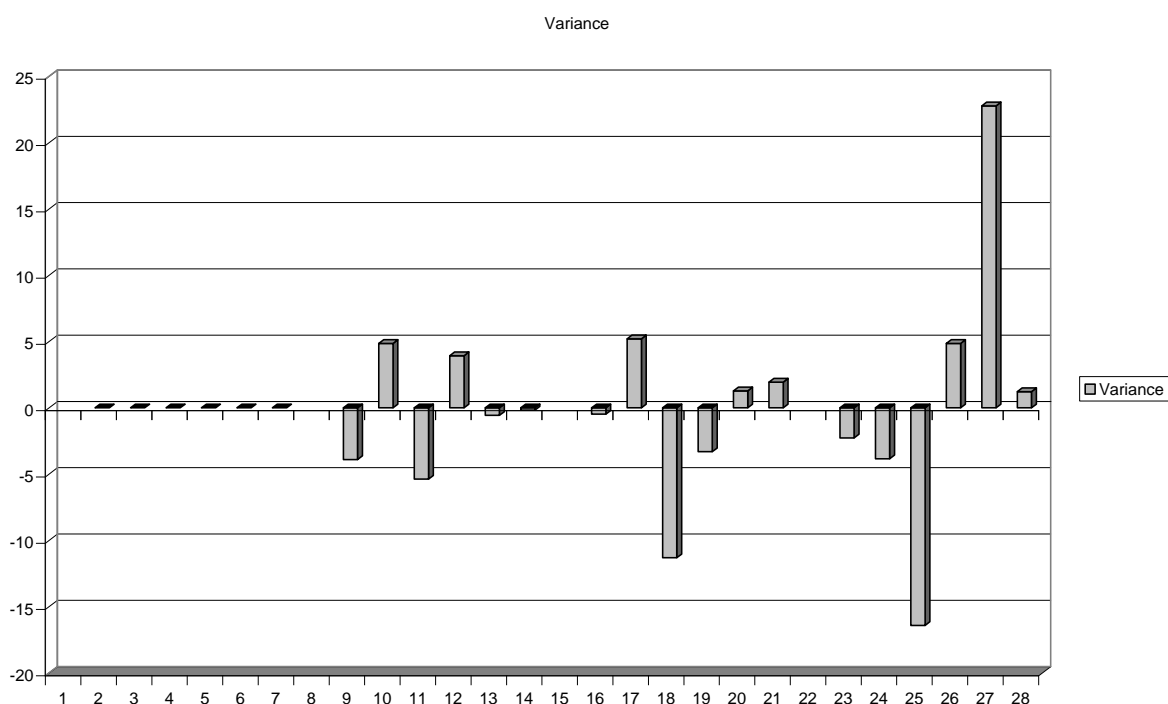
The function of statistics is to use known vectors in the data to discover or prove other vectors. In this case, the discrete binomial vector of marked / unmarked form has been analysed in relation to the discrete vectors of verb-type, tense and number to try to identify the effect of the continuous vector of idiom. However, it has been found that a vector of selfhood, possibly continuous, has prevented a strict numerical analysis. Untangling these two vectors is not simple and work remains to be done; but the principle, that markedness is not isolable purely on the basis of grammar, appears to be proved.

Figure 1 – December 2002 data

					Total	Unmarked %	Predicted	Variance
						Graph 1		Graph 2
<b>1</b>		<b>Me</b>	<b>Myself</b>					
2	I love	6075	7129		13204	46.00	46.00	0.00
3	I like	4255	3218		7473	56.93	56.93	0.00
4	I do not like	66	84		150	44.00	44.00	0.00
5	I don't like	227	881		1108	20.48	20.48	0.00
6	I dislike	16	87		103	15.53	15.53	0.00
7	I hate	1046	13363		14409	7.25	7.25	0.00
<b>8</b>		<b>Me</b>	<b>Myself</b>					
9	I loved	148	728		876	16.89	20.79	-3.90
10	I liked	145	329		474	30.59	25.73	4.86
11	I did not like	9	53		62	14.51	19.89	-5.38
12	I didn't like	74	487		561	13.19	9.26	3.93
13	I disliked	4	58		62	6.45	7.02	-0.57
14	I hated	139	4333		4472	3.10	3.28	-0.18
<b>15</b>		<b>Us</b>	<b>Ourselves</b>	<b>Ourself</b>				
16	We love	811	4737	146	5694	14.24	14.72	-0.48
17	We like	119	386	3	508	23.42	18.22	5.20
18	We do not like	1	35	0	36	2.77	14.08	-11.31
19	We don't like	6	178	1	185	3.24	6.55	-3.31
20	We dislike	2	30	0	32	6.25	4.97	1.28
21	We hate	27	603	3	633	4.26	2.32	1.94
<b>22</b>		<b>Us</b>	<b>Ourselves</b>	<b>Ourself</b>				
23	We loved	14	79	1	94	14.89	17.16	-2.27
24	We liked	4	19	0	23	17.39	21.23	-3.84
25	We did not like	0	0	0	0	0.00	16.41	-16.41
26	We didn't like	2	14	0	16	12.50	7.64	4.86
27	We disliked	2	5	0	7	28.57	5.79	22.78
28	We hated	2	49	0	51	3.92	2.70	1.22

Figure 2 – Unmarked forms as percentages of all forms



**Figure 3 – Variance between unmarked % and predicted values based on first data set (2 to 7)****Figure 4 – context-related data for *I hate me / myself* and *we like us / ourselves***

I hate me	28
I hate me because...	1
I hate me whenever...	1
	<b>30</b>
I hate myself	12
Emphatic (this is what I hate myself)	2
I hate myself because...	8
I hate myself when...	5
I hate myself for...	3
	<b>30</b>
We like us	19
we like us as we are	5
We like us best	3
We like us this way	2
We like us some Texas country	1
	<b>30</b>
We like ourselves (corporate)	14
We like ourselves (each other)	9
Emphatic (e.g. we only list sites that we like ourselves)	7
	<b>30</b>

## REFERENCES

- Aristar Dry, Helen and Anthony Rodrigues Aristar (1998) 'The Internet – an introduction'. In John Lawler & Helen Aristar Dry (eds) *Using Computers in Linguistics – a practical guide*, ch.2. London: Routledge.
- Bates, Elizabeth, Philip S Dale and Donna Thal (1995) 'Individual Differences and their Implications for Theories of Language Development'. In Paul Fletcher & Brian MacWhinney (eds) *The Handbook of Child Language*, ch.4. Oxford: Blackwell Publishers Ltd.
- Bruner, Jerome (1990) 'Autobiography and Self'. In *Acts of Meaning*, ch.4. Cambridge: Harvard University Press.
- Edwardes, Martin (2001) '6.6. The Component Grammar Survey'. In *Grammar and Language: a study of the ways in which grammar creates and is created by language* ms (MA dissertation) University of East London. Available at: [http://www.btinternet.com/~martin.edwardes/linguistics\\_ma\\_ch6.htm](http://www.btinternet.com/~martin.edwardes/linguistics_ma_ch6.htm)

- Giddens, Anthony (1991) 'Modernity and Self-Identity: tribulations of the self', in Adam Jaworski and Nikolas Coupland (eds) *The Discourse Reader*, ch.24. London: Routledge.
- Grice, Paul (1969) 'Utterer's Meaning and Intentions'. In *Studies in the Way of Words* (1989), ch5. Cambridge: Harvard University Press.
- Huang, Yan (2001) 'Anaphora, Syntax and Neo-Gricean Pragmatics'. Paper presented as part of *The Workshop on Pragmatics and Anaphora*, at the LAGB conference, Reading, UK, 6 September 2001.
- Lakoff, George; and Mark Johnson (1980) *Metaphors we Live By*, Chicago: University of Chicago Press.
- Maxwell, John T III; and Ronald M Kaplan (1995) 'The Interface between Phrasal and Functional Constraints'. In Mary Dalrymple, Ronald M Kaplan, John T Maxwell III & Annie Zaenen (eds) *Formal Issues in Lexical-Functional Grammar*, ch.15. Stanford: CSLI Publications.
- Peacocke, Christopher (1997) 'Holism'. In Bob Hale and Crispin Wright (eds) *A Companion to the Philosophy of Language*, ch.10. Oxford: Blackwell Publishers Ltd.
- Perry, John (1997) 'Reflexivity, Indexicality, and Names'. In Wolfgang Kunne, Albert Newen and Martin Anduschus (eds) *Direct Reference, Indexicality, and Propositional Attitudes*, ch.1. Stanford: CSLI Publications.
- Eddie Williams (2001) 'Unity, Diversity and Legitimation in Verb Phrase Semantics: football and fiction'. Paper presented at the BAAL conference, Reading, UK, 6-8 September 2001.
- Wray, Alison (2000) 'Holistic Utterances in Protolanguage: the link from primates to humans'. In Chris Knight, Michael Studdert-Kennedy and James R Hurford (eds) *The Evolutionary Emergence of Language - social function and the origins of linguistic form*, ch.17. Cambridge University Press, Cambridge, UK.
- Wray, Alison (2002) 'Dual Processing in Protolanguage: performance without competence'. In Alison Wray (ed.) *The Transition to Language*, ch.6. Oxford: Oxford University Press.
- Xue, Ping and Fred Popowich (2002) 'The Dual Status of Middle-Distance Reflexives'. *Journal of Linguistics* 38.1, pp.71-86. Cambridge: Cambridge University Press.
- Zribi-Hertz, Anne (2001) 'Reflexivity and Disjoint Reference: where does syntax begin?' Paper presented as part of *The Workshop on Pragmatics and Anaphora*, at the LAGB conference, Reading, UK, 6 September 2001.

*Martin Edwardes*

School of Social Sciences  
University of East London  
Barking Campus  
Longbridge Road  
Dagenham  
Essex RM8 2AS

[martin.edwardes@btopenworld.com](mailto:martin.edwardes@btopenworld.com)  
<http://www.btinternet.com/~martin.edwardes/>