# HALLIDAY'S PROBABILITY RESETTING AND METAPHOR USE: COMPARING A SPECIFIC GENRE TO GENERAL LANGUAGE

*Tony BERBER SARDINHA*
*(Catholic University of São Paulo, Brazil)*

ABSTRACT: *In this paper, I take a look at the extent to which the distribution of lexical metaphor matches the probability patterns proposed by Halliday. The research was motivated by the fact that lexical metaphor has long been ignored in SFL, and so there was no indication that the use of lexical metaphor would be predicted by patterns of distribution that were originally meant to refer to typical systemic categories (such as polarity). The research was based on the analysis of two corpora, one of conference calls and another of general language, both of Portuguese. The conference calls were read one by one and all of its metaphors were coded; then the same metaphorical lexis was searched for in the general language corpus and analyzed. After that, the frequency of metaphor use was calculated for each corpus, its distribution was described in both corpora, and this was compared to Halliday's original proposal. Finally, a comparison was carried out between the two corpora to see to what extent there was 'resetting', or shifts in probability patterns, as predicted by Halliday. The results indicated that metaphor use largely matched Halliday's probability profiles, and there was evidence of resetting, with the exception of cases where 'zero resetting' occurred. The paper closes with remarks arguing for Systemic Functional Linguistics to stop ignoring lexical metaphor, as it is a very important device for linguistic expression as well as link to conceptual structures in the mind.*

## 1. Probabilities in linguistic theory

Probabilistic approaches are not the dominant paradigm in linguistics; as Chambers (1995) (cited in Manning, 2003: 290) puts it, linguistics still suffers from a 'tradition of categoricity', that is, one in language is seen as made up of discrete categories that carry no information about the likelihood of it happening at all[i]. It was Chomsky who made the most scathing criticism of the probabilistic position, by arguing 'it must be recognized that the 'probability of a sentence' is an entirely useless one, under any known interpretation of this term (Chomsky, 1969:57) (apud Manning, 2003:289). This had the effect of killing off interest in probabilities for a long time, until the advent and dissemination of the

computer corpus made it inevitable that probabilities came to the fore once again.

The main proponent of a probabilistic framework for language is Michael Halliday (1991; 1992; 1993), who draws attention to the role of probabilities in shaping abstract linguistic systems. His general claims are that probabilities are essential properties of linguistic systems, and that seeking these probabilities in corpora and attaching them to the options available in the linguistic system is a way of enriching language description. We must bear in mind in the following discussion that Halliday does not refer to metaphor (lexical or grammatical) in his account of probabilistic systems, largely restricting his examples to functional systems such as polarity. However, there does not seem to be any impediments to our doing so. In fact, his ideas about language as probabilistic systems are general enough for us to apply them to issues related to metaphor use. It must be stressed that in this paper by metaphor is meant *lexical metaphor* and not *grammatical metaphor*.

Halliday began his work on probabilities in the linguistic system when he was preparing his grammar of Chinese, in the mid 1950's. Later, he calculated probabilities for English, taking samples of 2000 clauses from four different registers and counting grammatical features in them. His overall goal was this:

'What I hoped to do was to model each system not just as 'choose *a* or *b* or *c*', but as 'choose *a* or *b* or *c with a certain probability attached to each*. In other words, I was positing that an inherent property of any linguistic system is the relative probability of its terms.' (Halliday, 1993: 3)

In 1991, Halliday spent time at Cobuild, the large corpus linguistic lexicography project based in Birmingham, where he resumed his work on systemic probabilities, this time using large electronic corpora. This allowed him to derive the probabilities from larger amounts of data, which resulted in more robust figures. To his surprise, the probabilities taken from the large corpus resembled his previous ones from small non-electronic corpora. For polarity, the early probabilities were .9 positive and .1 negative, whereas the later ones were .8675 and .1325, respectively. For primary tense, they had been .5 present and .5 past, and then were .4955 and .5041.

His findings showed that there was a tendency for probabilities to be distributed either as .9 to .1 or as .5 to .5 (Halliday, 1993: 9). The former,

he called 'skew' distributions, and the latter, 'equiprobable'. Drawing on Shannon and Weaver's model of communication, Halliday proposed that skew systems are less informative, as there is less uncertainty in relation to users selecting the highly probable choice. Equiprobable systems, on the other hand, are more informative, as there is greater freedom of choice for users in selecting either one of the options available.

The fact that probabilities are part of linguistic systems alters the way we understand the concept of 'choice' in language use, as it does not mean 'free choice' any longer. As Halliday explains:

'[Language users] could choose to use negative more often than positive, just as they could choose to use *stroll* more often than *walk* – but they won't. … *grammatical* choices are governed by overall patterns of probability.' (Halliday, 1993 :3).

These overall patterns of probability vary according to a number of different situations. For example, probabilities attached to a given feature may be different for a specific register than they are for language as a whole. According to Halliday (1993: 24), register variation can be defined as 'the resetting of probabilities in the lexis and in the grammar.'

Probabilities also change over time, and this describes language change diachronically. The process of change is dynamic:

'each instance, that is, every piece of language that is actually spoken or written nudges the probabilities of the system, so that the system may change in the course of time.' (Kilpert, 2003: 183).

Halliday's work on probabilities was taken up by other systemicists, who investigated a range of different aspects of language (e.g. (Matthiessen, 2002; Nesbitt & Plum, 1988; Shimazumi, 1996) and their probabilities of occurrence.

Outside of Hallidayan linguistics, there is a wealth of research showing the potential of probabilities in enriching language description, teaching, and engineering. This is a result of the growth of corpus linguistic studies reporting frequency of use of linguistic items and structures (Berber Sardinha, 2004).

Perhaps the largest, best documented examples are Biber et al's (1999) 'Grammar of Written and Spoken English' and their companion Student Grammar (Biber, Conrad, & Leech, 2002), which describe in detail the

frequency of occurrence of grammatical structures across different major registers. Their explanations make it clear that the probability of structures happening in language varies according to the registers in which they occur. For instance, the odds of a 'do' negative contraction occurring in conversation is 100%, whereas in fiction it is around 70%, in news it is 60% and in academic writing it is below 10% (Biber et al., 2002: 242).

In linguistic theory, Bod, Hay and Jannedy (2003c) bring together investigations in several sub-fields of linguistics, including psycholinguistics, morphology, phonology, syntax and semantics. According to them:

'A wide variety of evidence suggests that language is probabilistic. In language comprehension and production, probabilities play a role in access, disambiguation, and generation. In learning, probability plays a role in segmentation and generalization. In phonology and morphology, probabilities play a role in acceptability judgments and alternations. And in syntax and semantics, probabilities play a role in the gradience of categories, syntactic well-formedness judgments, and interpretation. Moreover, probabilities play a key role in modelling language change and language variation.' (Bod, Hay, & Jannedy, 2003b: vii).

In the field of teaching English as a foreign language, the role of probabilities has also been recognized. McCarthy and Carter (2001), for instance, provide a probabilistic account of the use of *get*-passives and discuss it as such:

'We would argue that such probabilistic statements are in fact extremely useful in a pedagogical grammar; indeed, it is hard to envisage a proper description of the *get*-passive that would be pedagogically useful without including information for the learner about its overwhelming probability of occurrence in informal spoken contexts, with unfortunate events, and the unlikelihood of the occurrence of a typical *by*-agent phrase.' (McCarthy & Carter, 2001: 57).

Finally, in computational linguistics and in Natural Language Processing, probabilities play a central part in the development of software such as taggers and parsers. A probabilistic tagger (or a parser) is one that assigns tags based on probabilities drawn from a previously tagged corpus (eg. Beale, 1985; Garside & Leech, 1985), without the application of pre-defined rules about how parts of speech, for example, should be determined for unknown words (e.g. in 'Article

unknown_POS Noun', 'unknown_POS' would be tagged as 'adjective', unless other rules specify otherwise).

## 2. Metaphor

Three main kinds of metaphor are distinguished in contemporary metaphor theory: grammatical, conceptual and linguistic (Berber Sardinha, in print). Grammatical metaphor is a construct introduced by SFL to refer to the 'transferred' realizations in the lexicogrammar (Halliday, 1994 :342) which cause 'tension between wording and the meaning the wording represents' (Martin, 1999 :34), resulting in an 'association of the category meaning of one semantic category with another by replacing the original semantic category by a different semantic category in the lexicogrammar' (Lassen, 2003 :25). For example, the clause 'he published a book' is considered 'congruent' or non-metaphorical, because the process of publishing is realized by its typical category of the grammar, the verb and its complements. By contrast, 'the publication of his book' is considered 'incongruent' or 'metaphorical' because the process of publishing is now realized by a noun, whose congruent function is not to represent processes but things. As a result, the act of publishing a book has now become a thing through metaphor.

The second main type of metaphor is conceptual metaphor. This comes from a different tradition of linguistic inquiry associated with cognitive linguistics. A conceptual metaphor is one which 'brings together two domains that are distinct and somehow incongruous, but whose juxtaposition can be made sense of.' (Cameron, 2002, p.674). These two domains are termed 'source' and 'target' domains (Lakoff & Johnson, 1980). Take the following use of 'apostas' from our data:

(1) 'Investidores estrangeiros continuam aumentando suas apostas na Bolsa paulista.'
(Foreign investors continue to raise their bets on the Paulista Stock Exchange.)

In this example, two distinct domains are mapped together, namely investment (the target domain) and games/gambling (the source domain) (Koller, 2004: 68-69). As a result, we may detect a conceptual metaphor underlying this fragment as INVESTING MONEY IN THE STOCK EXCHANGE IS GAMBLING.

The third main type of metaphor is linguistic metaphor. This is referred to in SFL as 'lexical metaphor'. A linguistic metaphor is 'a stretch of language that creates the possibility of activating two distinct domains' (Cameron, 2003: 674). There are no hard and fast rules for determining how long such as stretch is, and so we have to determine this based on the presence of Topic and Vehicle terms. The Vehicle terms are 'the words that signal the incongruous domain' (Cameron, 2003: 674). In the previous example, the Vehicles are 'aumentando suas apostas'. It signals the source domain of gambling. The remaining words are the 'Topic', and they signal the target domain of investment. In this way, the linguistic metaphor turned out to be the whole sentence, but this is not necessarily true in all cases. Sometimes linguistic metaphors are shorter, and more than one of them may cluster together in one single sentence or, in the case of talk, in a single utterance or intonation unit.

In this paper, we will deal with conceptual and linguistic metaphor only. The reason for this is that we want to argue that the behavior of conceptual and linguistic metaphor in language can also be captured by the probability patterns suggested by Halliday, even though he was not originally referring to lexical metaphor. It must be stressed that Halliday never claimed that the probability patterns only applied to the major Systemic categories. Hence, there is no theoretical reason to suppose that lexical metaphor use would not fit into these patterns. However, this has never been demonstrated in the literature. If our data support this (and there is no reason to suppose it would not), then it might mean that lexical metaphor is not entirely different from other functional systems (such as polarity, transitivity, modality, and so on), and so there should be room in SFL to deal more often and directly with lexical metaphor as well.

## 3. Data and Method

The data for this study consisted of two corpora. The first is a register-specific corpus, containing 14 conference calls, totalling 85,438 tokens (5,194 types; henceforth, specialized corpus). These conference calls are phone calls hosted in Portuguese by a Brazilian investment bank between its board of directors and the general public, investors, shareholders and the press. The second corpus is a large, register-diversified corpus of Portuguese called 'Banco de Português', with nearly 240 million tokens (henceforth, general corpus).

For the analysis of word classes, the corpora were run through a part-of-speech tagger for Portuguese, namely QTAG (Mason, 1997), which had

been trained for Portuguese. This tagger is available online at http://lael.pucsp.br/corpora.

The specialized corpus was fully hand-annotated for metaphor Vehicles by the author. This consisted of reading each text and tagging Vehicles with a special code. There were three rounds of annotation, in order to ensure a consistent and thorough tagging. Upon completion of the hand-annotation phase, a total of 423 metaphor Vehicle word forms had been identified.

The general corpus was used as follows. For each of the 423 metaphor Vehicle word form found in corpus A, a concordance was run in the general corpus. A maximum of 100 randomly extracted concordance lines was allowed for each word form. Each concordance was then analyzed manually and the metaphorical uses of each word form were tagged.

A computer program was written by the author to compute the probabilities of each word form being a metaphor Vehicle in the specialized corpus, the general corpus and jointly in the two corpora.

## 4. Results

In both corpora, the probability of being a metaphor is skew, in Halliday's (1991) term. Using a one-decimal representation of the ratios gives us the classic skew distribution that Halliday demonstrated, that is, a .9 probability attached to metaphorical use versus a .1 probability for non-metaphorical use in the specialized corpus and .7 : .3 for the general corpus. This suggests that in both corpora, the metaphorical sense of the Vehicle word forms are dominant, but they are much more so in the specialized corpus.

According to Table 1 the majority of cases (76%) involved probabilities moving upward, that is, they were higher in the specialized corpus than in the general one, thus confirming the overall probabilities presented earlier. However, in 63 cases (15%), probabilities were not reset at all. And in 37 cases (9%), probabilities moved downward, that is, reset the other way round, being greater in the general corpus.

**Table 1.** Summary of probability resetting

| | |
|---|---|
| Upward resetting (general corpus < specialized corpus) | 323 |
| No resetting (general corpus = specialized corpus) | 63 |
| Downward resetting (general corpus > specialized corpus) | 37 |
| Total | 423 |

**Table 2.** Resetting breakdown by average probability and *hapax legomena* (once occurring) Vehicles

| Direction of resetting | Average probability of Vehicles | | *Hapax Legomena* among Vehicles | | | |
|---|---|---|---|---|---|---|
| | Corpus | | Corpus | | | |
| | Specialized | General | Specialized | | General | |
| Upward (General < Spec.) | .97 | .62 | 119 | 36.8% | 13 | 4.0% |
| No resetting (General = Spec.) | 1.00 | 1.00 | 33 | 52.4% | 0 | 0.0% |
| Downward (General > Spec.) | .48 | .77 | 12 | 32.4% | 0 | 0.0% |

Table 2 shows the quantitative profiles of Vehicles in terms of their probability resetting.

According to these numbers, words that had an upward shift in probability (General < Spec.) are normally rare metaphors both in general language (4% occurring only one) and in the conference calls

(37% occurring only once). Words that underwent no resetting are frequent ones whose metaphorical meaning is the dominant one in Portuguese, including the context of banking. Finally, words that had downward probability resetting (Spec. < General) tend to be general words with frequent metaphorical senses in the language, which are rarer in the conference calls.

Our findings suggest that probabilities of metaphor use are high or moderately high in general. Our results also show that there is a higher probability associated with a word form recurring as a Vehicle in a specialized corpus (.86) than in general language (.73). This is in accordance with Halliday's (1991; 1993) theory that suggests that registers reset the general probabilities of language. In our case, this resetting was in the direction of making probabilities stronger in the specialized register, but this does not need to be necessarily the case, at least in theory.

Metaphors form a skew distribution, of the kind that Halliday (1991) found for systems in the grammar of English such as polarity. These figures provide quantitative corroboration of the systematic nature of metaphor (Lakoff & Johnson, 1980). More specifically, these high probabilities in specialized language also reflect the metaphorical nature of terminology, as certain metaphors become the 'default' or 'unmarked' way of referring to a number of concepts and topics in the domain of market trading.

Probability resetting, although seen by Halliday (1991; 1993) as a natural consequence of register variation, revealed patterns that could not have been predicted from his theory. A closer look at the probabilities indicated that resetting did not occur across the board. In 9% of the cases, there was no resetting; Vehicles retained their general language probability. In addition, probabilities shifted in both directions, that is, being higher in either corpus in comparison with the other. This led us to differentiate upward, downward and zero resetting.

Metaphorical meanings seemed to vary with the direction of the resetting. In upward resetting, we found the bulk of the metaphors in the conference calls. This is also the most diverse grouping of Vehicles, including somewhat rare terminology ('locking a currency value', 'building something into a projection'), general business concepts ('fluctuation', 'restructuring', 'network') as well as colloquial, colorful language ('crystal ball', 'snapshot', 'holding fire').

Zero resetting is perhaps the most interesting case, as it was not predicted by Halliday's (1991; 1993) theory, at least explicitly. Among Vehicles in this category were well established, stable metaphors coming from the business domain (volatility, fluctuation) as well as words that seem to have migrated from specialized domains to more general use (such as goals, niche). It is likely that the press has had a large role to play in popularizing these metaphors, and the fact that our general language corpus has considerable amounts of newspaper text may have influenced our results in this direction.

## 5. Conclusion

This study has shown that lexical metaphor is very often present in language, both in specialized registers and in language in general.

The probability profiles of metaphor Vehicles followed the basic patterns predicted by Halliday, with a .9 : .1 profile in favor of metaphor use for the specialized corpus probability and a .7 : .3 profile for general language. This suggests that lexical metaphor behaves broadly as any other functional system in the language, even though it lies outside the main scope of the SFL theory. This might be taken as meaning that lexical metaphor is similar to other systems in the Systemic framework, at least as far its distribution, and so there would be no reason to ignore it in Systemic Functional Linguistics. By focusing on grammatical metaphor only, SFL overlooks a major element of linguistic expression, which relates to language and the mind. This does not mean of course playing down the role of grammatical metaphor, but rather arguing the need for Systemics to look more closely at lexical metaphors as well.

Understanding how metaphor works in language and in the mind is an important endeavor, one that ultimately helps us understand how human beings think, construe and exchange meaning. It is hoped that this study may have made a contribution to this by bringing together approaches to metaphor that are usually distant from each other.

**REFERENCES**


BEALE, A. D. 1985. *A probabilistic approach to grammatical analysis of written English by computer*. Paper presented at the Second Conference of the European Chapter of the Association for Computational Linguistics, Geneva, Switzerland.

BERBER SARDINHA, T. 2004. *Lingüística de Corpus.* São Paulo: Manole.

_____. in print. *Metáfora.* São Paulo: Parábola.

BIBER, D., S. CONRAD, & G. LEECH. 2002. *Student Grammar of Spoken and Written English.* Harlow: Longman.

BIBER, D., S. JOHANSSON, G. LEECH, S. CONRAD, & E. FINEGAN. 1999. *Longman Grammar of Spoken and Written English.* Harlow: Longman.

BOD, R., J. HAY, & S. JANNEDY. 2003a. Introduction. In R. Bod, J. Hay & S. Jannedy. Eds. *Probabilistic Linguistics*. Cambridge, Mass.: The MIT Press.

_____. 2003b. Preface. In R. Bod, J. Hay & S. Jannedy. Eds. *Probabilistic Linguistics*. Cambridge, Mass.: The MIT Press.

_____. 2003c. *Probabilistic Linguistics.* Cambridge, Mass: MIT Press.

CAMERON, L. 2002. Metaphors in the learning of science: A discourse focus. *British Educational Research Journal, **28***(5): 673-688.

_____. 2003. *Metaphor in Educational Discourse.* London: Continuum.

CHAMBERS, J. K. 1995. *Sociolinguistic theory : linguistic variation and its social significance.* (2nd ed.) Oxford, UK ; Malden, MA: Blackwell Pub.

CHOMSKY, N. 1969. Quine's empirical assumptions. In D. Davidson, W. V. Quine & J. Hintikka. Eds. *Words and objections. Essays on the work of W.V. Quine*. Dordrecht: D. Reidel.

GARSIDE, R., & F. LEECH. 1985. *A probabilistic parser*. Paper presented at the Second Conference of the European Chapter of the Association for Computational Linguistics, Geneva, Switzerland.

HALLIDAY, M. A. K. 1991. Corpus studies and probabilistic grammar. In K. Aijmer & B. Altenberg. Eds. *English corpus linguistics: Studies in honour of Jan Svartvik*. London: Longman.

_____. 1992. Language as system and language as instance: The corpus as a theoretical construct. In J. Svartvik. Ed. *Directions in corpus linguistics - Proceedings of the Nobel Symposium 82, Stockholm 4-8 August 1991*. Berlin, New York: De Gruyter.

_____. 1993. Quantitative studies and probabilities in grammar. In M. Hoey. Ed. *Data, description, discourse*. London: HarperCollins.

_____. 1994. *An Introduction to Functional Grammar*. London: Edward Arnold.

KILPERT, D. 2003. Getting the full picture: A reflection on the work of M.A.K. Halliday. *Language Sciences, **25***: 159-209.

KOLLER, V. 2004. *Metaphor and Gender in Business Media Discourse: A Critical Cognitive Study*. Basingstoke: Palgrave Macmillan.

LAKOFF, G., & M. JOHNSON. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

LASSEN, I. 2003. *Accessibility and Acceptability in Technical Manuals. A Survey of Style and Grammatical Metaphor. .* Amsterdam/Philadelphia: John Benjamins.

MANNING, C. D. 2003. Probabilistic syntax. In R. Bod, J. Hay & S. Jannedy. Eds. *Probabilistic Linguistics*. Cambridge, Mass.: The MIT Press.

MARTIN, J. R. 1999. Modelling context: A crooked path of progress in contextual linguistics. In M. Ghadessy. Ed. *Text and Context in Functional Linguistics*. Amsterdam/Philadelphia: John Benjamins.

MASON, O. 1997. QTAG-A Portable Probabilistic Tagger (Version 1). Birmingham: University of Birmingham.

MATTHIESSEN, C. M. I. M. 2002. Systemic functional profiles of system and text: investigations based on texts, text archives and corpora, *International Systemic Functional Congress*. Liverpool, UK.

MCCARTHY, M. J., & R. CARTER. 2001. Ten criteria for a spoken grammar. In E. Hinkel & S. Fotos. Eds. *New Perspectives on Grammar Teaching in Second Language Classrooms*. Mahwah, NJ: Lawrence Earlbaum.

NESBITT, C., & G. PLUM. 1988. Probabilities in a systemic grammar: The clause complex in English. In R. P. Fawcett & D. Young. Eds. *New Developments in Systemic Linguistics*. London: Pinter.

SHIMAZUMI, M. 1996. *The Knower and the Informant in Institutional Talk: A Transitivity Perspective.* AELSU, University of Liverpool, United Kingdom.

[i] As Bod (2003a) argues, categorical and probabilistic approaches are not incompatible, though. In fact, they are simply two perspectives on the same phenomenon. Categorical approaches focus on the endpoints of distributions, whereas probabilistic ones include the points in between as well.