

# *Polifonia de discursos: análise computacional de um corpus literário*

Sonia Syngier

## **1. Introdução**

Este artigo nasce de uma preocupação com o ensino da literatura. Aqui parte-se do pressuposto de que a sala-de-aula de literatura é um *locus* de emancipação em que o aluno exercita a liberdade de ler e interpretar. Para tanto, demonstra-se como o computador pode ser um instrumento adequado para propiciar um ambiente em que os alunos pensam sobre a uso estético da linguagem e produzem avaliações bem fundamentadas sobre suas leituras. Justamente porque se sentem aptos a trabalhar a linguagem crítica, ficam menos vulneráveis ao poder da crítica impressa. Ao encontrar suas próprias respostas, os alunos podem, então, desafiar leituras tradicionais, buscar o implícito e não se apoiar no que já foi dito. O olhar pode circular pelas entrelinhas do texto e o leitor passar a fazer inferências com base na linguagem utilizada para construir o texto.

Perspectivas relativistas e teorias voltadas para a recepção do leitor ao texto sustentam que a percepção de padrões é um evento cultural que depende de uma convenção de relações construídas socialmente. Portanto, a questão da variabilidade das interpretações de acordo com grupos e através das gerações já foi amplamente discutida. Cabe lembrar, no entanto, que as propriedades textuais podem apresentar restrições às interpretações e justamente são essas restrições que respondem em muito pela qualidade do texto. Eco, por exemplo, afirma que “qualquer interpretação de um determinado trecho pode ser aceita se for confirmada, e deve ser rejeitada se for desafiada, por outro trecho do mesmo texto. Nesse sentido, a coerência textual interna controla as vontades inconsoláveis do leitor” (1992: 65, minha tradução)<sup>1</sup>.

Trata-se aqui de uma postulação muito interessante para a área de educação literária. Os alunos podem não ser leitores experientes e sofisticados que têm acesso a uma rica e complexa rede de referências. No entanto, esses mesmos alunos inexperientes podem lançar novas luzes sobre o texto e contribuir para a discussão se puderem justificar suas argumentações a partir das escolhas lingüísticas de um determinado texto.

## **2. Valor: uma nota**

Atribuir valor literário de forma definitiva e permanente é uma impossibilidade. Pode-se sugerir, levantar possibilidades, persuadir, mas as respostas a todas essas questões devem permanecer de forma provisória. A avaliação literária se caracteriza por ser provisória, assim como a questão do gosto. Uma das razões pode ser encontrada no envolvimento da crítica com fatores sociais e culturais. Con sidera-se literário um texto que uma determinada

comunidade busca preservar. Dito isso, não se invalida a possibilidade de se

estabelecer alguns critérios, mesmo que em caráter provisório. Nesse trabalho, parte-se do princípio de que a linguagem de um texto literário é construída de tal forma que coloca em foco valores de uma determinada comunidade de leitores.

### 3. Critérios para avaliação: previsibilidade e função

Como apontado na seção anterior, decidir sobre que textos devem ser valorizados não é uma questão fácil. Qualidade depende em grande parte de um acordo cultural. Neste trabalho, assume-se que qualidade se apoia pelo menos em dois critérios: previsibilidade e função. Segue a explicação:

A psicologia cognitiva e a teoria da informação sugerem que o que chama a atenção de um leitor a um determinado evento é a sua imprevisibilidade (cf. Bruner et al., 1951; Luria, 1976). Por exemplo, as manchetes dos jornais costumam quebrar a rotina dos eventos diários trazendo sempre algo de novo. Buscam fazer com que o leitor reflita sobre a novidade, compare-a com um quadro de expectativas e, uma vez dando-se conta da diferença, insira-a no seu repertório como um novo item. Gombrich (1986), por exemplo, mostra como um *outdoor* chama a atenção do espectador devido a sua improbabilidade, prolongando, desta forma, o processo de interpretação. A acomodação se segue e esse novo quadro se torna parte integrante da experiência vivida.

A imprevisibilidade já foi descrita como resultante da interrupção de uma rotina. Schmidt (1989) afirma que os seres humanos agem de acordo com sua rotina filogenética. Operam indutiva e previsivelmente. Esperam que o que lhes aconteceu uma vez acontecerá novamente. A repetição, portanto, é um aspecto inerente de qualquer sistema vivo. Nesse sentido, a novidade vai depender do grau em que a rotina das estruturas repetitivas são quebradas. Em termos lingüísticos, um texto se torna mais informativo quando é menos previsível. Com base no que ocorre textualmente, o leitor projeta suas expectativas ... e as têm negadas ou confirmadas. O texto pode também oferecer alternativas de escolha, resultando, então, em ambigüidade. Isso quer dizer que quanto mais imprevisível a linguagem de um determinado texto, mais literário ele o é?

Uma vez que a previsibilidade está associada à frequência de ocorrências (Sinclair, 1982; Willis, 1990), a possibilidade de se saber *a priori* que uma certa palavra, frase, padrão, etc. irá acontecer tem uma importante influência na determinação do valor de um texto. Segue-se daí que o desenvolvimento de pesquisas empíricas utilizando a lingüística de *corpus* pode ajudar sobremaneira.

No entanto, a linguagem de um texto não é fonte exclusiva de atribuição de valor estético. Em última instância, a função de um determinado texto é atribuída pelo leitor, que decidirá que direção dar à leitura. Vários textos viajam através da história e através de contextos, sem perder seu valor. Como pressuposto, assume-se nesse artigo que os textos que atenderam ao critério de (im)previsibilidade e função sem perder seu valor através dos tempos e dos lugares podem garantir seu *locus* no cânone.

### 4. O papel da análise de corpus

A análise de *corpus* não é um novo ramo da lingüística como a psicolingüística

ou a sociolinguística. Trata-se de uma metodologia através da qual usa-se o computador para organizar um vasto número de trechos escritos ou falados, permitindo a percepção de padrões que não seriam notados de outra forma. Geoffrey Leech<sup>2</sup> data o primeiro uso do termo em 1984. Segundo ele, ‘corpus linguistics’ só aparecia vez por outra até a publicação do livro *Corpus Linguistics: Recent developments in the use of computer corpora in English language research*, edited by Aarts & Meijs”. Define-se: um *corpus* é uma coleção de trechos linguagem selecionados e ordenados de acordo com critérios lingüísticos claros para ser usada como amostra de linguagem. Para Sinclair (1991: 17), “Um dos principais usos do *corpus* é para identificar o que é central e típico na linguagem”. Neste caso, se os escritores tendem justamente a inovar, cada texto literário deveria ser considerado único. Qual o papel, então, da análise de *corpus* na literatura?

A questão da novidade é muito complexa. Um texto pode ser considerado literário devido a uma série de condições e valores e não somente porque sua linguagem é criativa. *Mutatis mutandi*, nem todo uso criativo de linguagem é literário. Veja, por exemplo, o caso das propagandas e das piadas. Como qualquer instância de uso de linguagem, a criatividade depende de interpretação. Assim, quanto menos esforço para se interpretar, menos criativo o texto. Neste ponto, a noção de reversão semântica proposta por Sinclair (1997) pode ser esclarecedora. Ele explica que um texto depende de dois tipos de significado: o significado exato da palavra conforme determinado pelo léxico (ou pela frequência de seu uso) e o significado produzido pelo contexto verbal em que a palavra se insere. A linguagem não-criativa se constrói a partir do acúmulo de cada unidade sucessiva de significado da forma esperada, tornando o texto altamente previsível: espera-se por dois significados – aquele criado pelo contexto e aquele criado por cada item individualmente. Diferentemente, em um uso mais criativo (metáforas, ironia, etc.), o significado gerado pelo contexto encontra-se distante ou até mesmo entra em conflito com o significado do item, forçando o leitor/endereçado a se utilizar de suas habilidades interpretativas. Sinclair (1996: 4) assim define: “Sempre que o significado surge predominantemente do contexto verbal ao invés do termo em si, considera-se uma instância de *reversão semântica*”. Reversão semântica, portanto, é típica de criatividade lingüística.

A lingüística de *corpus* pode mostrar o contexto de certos itens lexicais utilizados por um determinado autor e, desta forma, permitir afirmações fundamentadas sobre possíveis interpretações. Pode-se, também, comparar o uso que um autor faz de certas palavras a um conjunto de textos contemporâneos demonstrando, assim, o grau de criatividade. Textos do século XX não oferecem problemas nessa direção, já que se tem acesso a uma variedade de materiais escritos. Na análise abaixo, apesar do acesso a um *corpus* do século XVII (o Corpus de Helsinki), a linguagem do tipo de textos que constituem esse *corpus* não oferece cobertura suficiente, levando a possíveis problemas de interpretação. No caso de escritores de épocas em que não havia tantos registros de linguagem escrita e falada como hoje em dia, é importante sempre lembrar que o *corpus* se

limita à produção do autor e periféricamente à de seus contemporâneos. Bill Louw<sup>3</sup> sugere, sempre que possível, a construção de um *corpus* da leitura de um determinado autor, promovendo, dessa forma, um *corpus* intertextual.

Apesar dessas limitações, um texto literário pode ser comparado empiricamente a uma norma. Se essa norma é uma coleção de textos do mesmo período ou região ou se se trata da coletânea das obras de um determinado autor, pode-se usar a metodologia de lingüística de *corpus* para interpretações mais demonstráveis e verificáveis. No caso da análise abaixo (inglês elizabetano/jacobino), não foi possível o acesso a uma vasta coleção de textos sob forma eletrônica. O *corpus* foi, então, composto pela obra completa de Shakespeare, o *Corpus* de Helsinki, organizado diacronicamente, e trechos das *Crônicas* de Holinshed.<sup>4</sup>

## 5. A análise

Nessa seção será demonstrada uma análise realizada com *Macbeth* com o intuito de verificar se uma abordagem de lingüística computacional ao um texto literário não contemporâneo pode trazer novas evidências e ajudar a validar, privilegiar ou refutar certas interpretações. A pesquisa obedece a paradigmas qualitativos e quantitativos, já que uma vez produzidas as listas de frequência, alguns itens são analisados individualmente, observando-se seu comportamento em determinado contexto.

### a. Listas de frequências

Utilizando-se a ferramenta MicroConcord ao *corpus* de Shakespeare compilado pela Universidade de Birmingham, produziu-se uma lista das palavras mais frequentes tanto em *Macbeth* (**MAC**) como no *corpus* das tragédias de Shakespeare (**STC**)<sup>5</sup>. O objetivo aqui foi verificar que item lexical chamava a atenção. É fato que essa decisão, em última instância, é subjetiva, mas não invalida a análise. Para se estabelecer significância avaliativa, a investigação seguiu uma orientação pragmática em que a prioridade é dada ao efeito que um determinado texto tem sobre o leitor. Trata-se de uma perspectiva que está de acordo com teorias que assumem a existência de estruturas que podem não estar aparentes de imediato para o leitor mas respondem pelo efeito literário. Há uma possibilidade de esse efeito ser causado pela frequência do uso de certos itens, pela imprevisibilidade de certas ocorrências, ou mesmo pela ausência de alguns itens, ou seja, daquilo que não foi selecionado.

Com base em afirmações da crítica tradicional de que *Macbeth* estava voltado para o futuro (“a man of ‘shoulds’ and ‘woulds’”, nas palavras da Profª. Aila Gomes em 1969), observou-se primeiramente as formas verbais mais frequentes e seu uso na peça.. Nenhum aparato sofisticado foi utilizado: um computador 486 e o programa *MicroCon cord*. O *corpus* usado como base: a

forma eletrônica das obras completas de Shakespeare publicada pela Oxford University Press. A Tabela 1 abaixo mostra a lista de frequência dos verbos:

MAC (1 7.830 itens)			STC (1 01 .352 itens)	
	Ocorrências	%	Ocorrências	%
<b>Time</b>	46	0.25	146	0.14
King	36	0.20	197	0.19
Lord	35	0.19	448	0.44
Man	30	0.16	189	0.18
Sir	29	0.16	259	0.25
<b>Sleep</b>	26	0.14	54	0.05
<b>Blood</b>	26	0.14	68	0.06
<b>Son</b>	26	0.14	73	0.07
<b>Thane</b>	25	0.14	*não incluído nas 400 ocorrências mais freqüentes	
Heart	22	0.12	122	0.12

Tabela 1: Verbos

Uma comparação entre a lista de freqüência dos verbos em toda a obra de Shakespeare (**STC**) e em *Macbeth* (**MAC**) não indicou nada que distinguisse essa peça das outras tragédias. Verbos como **will** e **shall** que poderiam ser relevantes para o tema do tempo não são mais freqüentes em *Macbeth*. Sempre se pode argumentar que a noção de tempo pode ser apresentada sob outras formas que não necessariamente verbos. Pode haver outras marcas de tempo aguardando investigação. A presente análise busca retirar a possibilidade desse tema ter ocorrido devido aos verbos. Por outro lado, a análise levanta a questão da palavra **enter**, não considerada anteriormente, mas significativamente mais freqüente em *Macbeth*. Utilizada principalmente nas direções de palco, essa palavra aponta para o dinamismo da tragédia. Há muito mais movimentação dos atores, principalmente por parte de Macbeth. A Tabela 2 mostra os verbos mais freqüentes em **MAC** comparados aos do *corpus* de tragédias shakespereanas (**STC**):

	MAC	STC
1	Is	Is
2	Be	Be
3	Have	Have
4	Do	Do
5	Enter	Will
6	Will	Are
7	Shall	Shall
8	Come	Come
9	Hath	Enter
10	Was	Let
11	Would	Would
12	Make	Am

Tabela 2: Lista comparativa da ordem de freqüência dos verbos

A Tabela 2 confirma a relevância de “enter” e mostra uma diferença significativa entre “make” em **MAC** e “am” na **STC**.

Uma vez examinados os verbos, passou-se a observar os substantivos. A Tabela 3 mostra a lista de freqüência:

MAC (1 7.830 itens)			STC (1 01 .352 itens)	
	Ocorrências	%	Ocorrências	%
<b>Time</b>	46	0.25	146	0.14
King	36	0.20	197	0.19
Lord	35	0.19	448	0.44
Man	30	0.16	189	0.18
Sir	29	0.16	259	0.25
<b>Sleep</b>	26	0.14	54	0.05
<b>Blood</b>	26	0.14	68	0.06
<b>Son</b>	26	0.14	73	0.07
<b>Thane</b>	25	0.14	*não incluído nas 400 ocorrências mais freqüentes	
Heart	22	0.12	122	0.12

Tabela 3: Lista de freqüência de substantivos

A ordem em que essa freqüência aparece é a seguinte:

	MAC	STC
1	<b>Time</b>	Lord
2	King	Sir
3	Lord	Love
4	Man	King
5	Sir	Man
6	<b>Sleep</b>	Time
7	<b>Blood</b>	Father
8	<b>Son</b>	Heart
9	<b>Thane</b>	Night
10	Heart	Heaven

Tabela 4: Lista comparativa da freqüência de substantivos

A comparação entre os substantivos mais freqüentes em **MAC** e na **STC** resulta no aparecimento de cinco palavras muito interessantes. Como **time**, **sleep**, **blood**, **son**, **thane** não são os mais freqüentes na **STC**, essas palavras poderiam estar mostrando algo bem específico de *Macbeth*.

**b. Contextualizando os itens listados**

Dando prosseguimento à análise, produziu-se uma concordância das palavras para verificar se o padrão lingüístico em que elas ocorriam poderia levar a alguma interpretação.

· **Time**

Essa palavras é duas vezes mais freqüente em *Macbeth* (46 casos, 0.257%) do que em *Othello* (30 casos, 0.095%), *King Lear* (27 casos, 0.087%), e *Romeo*

*and Juliet* (30 casos, 0.103%). Comparada a *Macbeth*, ela é 40% menos freqüente em *Hamlet* (47 casos, 0.127%). Em *Hamlet*, no entanto, a ordem de freqüência das palavras (cf. Tabela 4 acima) é a seguinte: Lord (225), King (210), Sir (76), Love (70), Father (69), Man (60), Time (47), Heaven (44), Night (37), Heart (30). Essa ordem não só indica que *time* não é a palavra mais freqüente na peça

como também ilustra o conflito em que se encontra protagonista, que se vê entre a lealdade ao seu rei e a necessidade de levar adiante a vingança exigida pelo fantasma de seu pai. Em termos de qualidade textual, se forem comparados os usos que Shakespeare e Holinshed fazem da palavra, pode-se notar que a maioria dos 19 casos em Holinshed são altamente previsíveis e exigem muito pouco esforço interpretativo. Aqui, *time* é usado principalmente em posições de advérbio, precedido por *at*, *in*, *for a long* ou colocações com itens previsíveis como *\* and place*, *tarie a \**. No texto shakespeariano, não só *time* é invocado e personificado, como ocorre com uma variedade de substantivos muito diferentes (*seeds of \**; *master of his \**, *expense of \**, *bank and shoal of \**, *spy o' th' \**, *gaze o' th' \**), de adjetivos (*blessed \**, *best \**, *olden \**, *recorded*, *woeful \**), e de verbos (*mock the \**, *fill up the \**, *rue the \**, *beguile the \**, *grant the \**), que demonstram o uso criativo que Shakespeare faz da palavra (vide Apêndice I).

### • Thane

A contextualização de **thane** mostra os títulos dos personagens (Thane of Cawdor; Thane of Fife; Thane of Glamis; etc.). Em 8 casos, integram frases nominais de títulos honoríficos, em 6 casos constroem-se em torno de semântica prosódica positiva. Somente em 3 casos, apresentam conotação negativa. 8 casos foram considerados gerais e inespecíficos. Se forem somados os 7 casos de títulos honoríficos aos 6 casos em que **thane** apresenta conotação positiva, pode-se entender a perspectiva de Macbeth, que vê os títulos honoríficos de forma positiva, como um objeto de seu desejo. Interessante observar-se que dos 25 casos, em 8 há pronomes em sua vizinhança (“I am”; “me”; “to me”; “my”; “our”; “thee”; “thy”; “thine”) e em 5 casos aparece a palavra “worthy”, que contribui para a ten são dramática criada pelo desejo de Macbeth e para a qualidade do texto de Shakespeare. Em Holinshed, isso não ocorre. Não há qualquer implicação nas atribuições de títulos. Isso fica mais claro ao se comparar “All haile, Makbeth, thane of Glammis!” (**HOL**), onde o título vem imediatamente após o nome, como uma conseqüência lógica e natural, ao impacto de “All hail, Macbeth! Hail to thee, Thane of Glamis.”(**MAC**), onde o “thee” se coloca como o centro de referência, uma ponte entre Macbeth e o que ele virá a se tornar.

### • Son

Uma concordância dessa palavra mostra que em 15 casos ela ocorre em direções de palco com relação ao filho de MacDuff. Como não se trata de texto falado, pode-se retirar essas ocorrências, restando, então 9, 5 das quais

trazem o significado de descendência ou prole. Não se trata de um achado muito intrigante se tomado *per se*. Mas se colocado próximo ao tratamento dado à palavra **woman** (vide Tabela 8 e análise abaixo), o quadro muda de figura. Muitos dos personagens têm filhos na peça (Duncan tem 2, Banquo, MacDuff, Siward têm um cada). Foi inclusive especulada na crítica literária tradicional a possibilidade de Macbeth ter tido filhos (cf. Knights, 1933, entre

outros, mas como essa especulação não encontra confirmação textual, ela foi descartada.). Passa-se, então à análise das palavras **blood** e **sleep**.

**c. Análise de baixo para cima: 3 itens**

· **blood**

*Macbeth* é definitivamente a tragédia mais sangrenta de Shakespeare. A Tabela 5 apresenta a freqüência de **blood** comparada a *Hamlet* (**HAM**), *Othello* (**OTH**) e *King Lear* (**LRQ**)<sup>6</sup>, o *corpus* de Helsinki constituído de documentos da época de Shakespeare (**HCE2**) e o *corpus* de Holinshed (**HOL**), montado especificamente para esse estudo. A comparação dessa palavra conforme usada no *corpus* das tragédias (0.06%) e em *Macbeth* revela que ela é 2,08 vezes mais freqüente nessa peça:

MAC	HAM	OTH	LRQ	HCE2	HOL	STC
0.14%	0.05%	0.04%	0.04%	0.01%	0.01%	0.06%

Tabela 5: Ocorrência de “blood”

Em termos de coligação (ou a posição sintática em que o item ocorre), há pouco a se dizer sobre o padrão de ocorrência. Shakespeare parece preferir **blood** em posições remáticas, geralmente precedida por uma preposição (*with* \*, *in* \*). Há poucas instâncias de **blood** em posição temática. No entanto, a preferência semântica é bastante interessante. **Blood** é usado em relação a um babuíno, uma porca, e um morcego, o que contribui para a atmosfera exótica de bruxaria, mas não é isso que distingue o uso desse termos das outras peças. Nestas, o significado de **blood** é metafórico. Em *Macbeth*, há 15 usos literais e 11 metafóricos. Desta forma, o impacto visual e a violência da peça se tornam mais óbvios. A Tabela 6 indica como a razão entre o uso literal e o metafórico é maior em *Macbeth*:

	Metafórico	Literal	Total	L/M ratio
<b>MAC</b>	10	15	25	1.5
<b>HOL</b>	16	8	24	0.5
<b>HCE</b>	16	6	22*	0.37
<b>HML</b>	17	2	19	0.11
<b>LRQ</b>	12	4	16	0.33
<b>OTH</b>	10	2	12	0.2

\* a maioria em documentos religiosos

Tabela 6: Significado de “blood”

Com relação às colocações, em **MAC**, a maioria ocorre antes do nóculo<sup>7</sup>. Há preferência semântica para profusão, principalmente nas posições N-3 e N-2 (all, ocean, much, fountain, pour). Por exemplo, pode-se contrastar “gouts of blood” (**MAC**) a “that drop of blood” (**HAM**). Em **HOL**, 5 das 8 ocorrências literais traduzem uma idéia de profusão (“cakes of blood”, “beraied with blood”, “great effusion of blood”, “blood ran about”). Mesmo assim, não se compara à preferência pelo uso literal dessa palavra em **MAC**. Acresce que a associação de “blood” e “gold” em “His silver skin lac’d with his golden blood”<sup>8</sup> e

“gild the faces of the grooms” em **MAC** produz uma preferência semântica por riqueza, por algo valioso, o que não ocorre em **HCE**, **LRQ**, **HAM**, **OTH** ou **HOL**. O fato de Shakespeare preferir o significado literal ao metafórico não diminui a qualidade literária da peça. Se for comparado o uso que ele faz de “blood” e como Holin shed o faz, nota-se que o uso de Holin shed, embora mais metafórico, é altamente previsível. Daí a complexidade e riqueza do texto shakespeareano. Em Holinshed, a maioria dos casos se relacionam à descendência e à religião – como de se esperar. Em Shakespeare, a palavra é usada em aliterações (“blood-baltered Banquo”), em repetições (“blood will have blood”), com uma variedade de advérbios de intensidade, em colocações incomuns, e em situações muito concretas, apelando para os sentidos do leitor/ouvinte (“smell of the blood”, “thy blood is cold”). Esta variedade de usos suscita uma diversidade de experiências que transcendem o tempo e podem aqui indicar o grau de qualidade do texto (cf. Apêndice II). Como van Peer afirma (1997: 223), “ao facilitar o processo de engajamento da imaginação, [a literatura] proporciona aos leitores a oportunidade de experiência emocional, contemplação intensa, auto-avaliação, e uma possível reorientação para o mundo do dia-a-dia”.

· **sleep**

Outra palavra que leva a um questionamento das posições tradicionais em relação à peça é “sleep”. A Tabela 7, abaixo mostra sua ocorrência em diferentes peças.

<b>MAC</b>	<b>HAM</b>	<b>OTH</b>	<b>LRQ</b>	<b>HCE</b>	<b>HOL</b>
0.14%	0.03%	0.01%	0.02%	0.0%	0.004%

Tabela 7: Freqüência de “sleep”

**Sleep** é 2.8 vezes mais freqüente em *Macbeth* (0.14%) do que no *corpus* de tragédias (0.05%). Além disso, é mais comum como um substantivo em *Macbeth* do que nas outras peças (**MAC**: 16 substantivos/9 verbos; **HAM**: 4 substantivos/ 8 verbos; **LRQ**: 4 substantivos/4 verbos; **OTH**: 3 substantivos/3 verbos)<sup>9</sup>. Esta evidência não só marca a qualidade do texto shakespeareano mas também dá uma nova dimensão à cena de sonambulismo, ao mundo de sonhos e à ambiguidade de *Macbeth*.

Já Holinshed utiliza colocações triviais, como “passed that night without anie sleepe comming in his eies”; “could not sleepe in the night time”; keepe

him still waking from sleepe”; “as he lay sleeping”. O mesmo se aplica à sua descrição dos guardas (“awaked them out of their droonken sleepe”). Se essa descrição for comparada à imagística animalesca usada por Shakespeare (“when in swinish sleep/Their drenched natures be”), pode-se demonstrar o valor literário da peça também através da polissemia que esse uso cria. “Sleep” é um conceito, não um processo. É a “the season of all natures”. Shakespeare usa **sleep** de forma mais complexa do que Holinshed. Há dois tipos diferentes de sono na peça: o que apresenta preferência semântica negativa, como “Curtained sleep”, “thralls of sleep”, “swinish sleep”, “equivocates him in a sleep”, o que faz o

guarda gritar “Murder” dormindo, e o que apresenta preferência semântica positiva, como “innocent sleep”, “sleep in spite of thunder”, “the benefit of sleep”, aquele que faz o guarda “laugh in his sleep”. Reinstala-se “foul is fair and fair is foul”. O sono pode ser bom ou mau. Acrescenta à atmosfera de incerteza da peça e apóia o tema de ambigüidade. Essa evidência vai contra a afirmação de Knight de que “sleep [is]... the gentle nurse of life ... ‘sleep’ is twined with ‘feasting’. Both are creative, restorative, forces of nature. So Macbeth and his Queen are reft of both during the play’s action” (1968: 148). A cena de sonambulismo se transforma em mais uma evidência da forte ligação entre os dois protagonistas principais. Enquanto um não consegue dormir, o outro cai num sono contínuo. O casal deixa de compartilhar uma vida documentada através de cartas e diálogos trocados e passam a viver em mundos distintos.

#### · **Fear**

A produção desta lista também pode ajudar a questionar interpretações tradicionais sobre a peça. Por exemplo, Knight (1968: 139) aponta para a frequência da palavra **fear** para afirmar que o medo está no coração dessa peça. Todos temem. Uma lista de frequência de fato mostra que **fear** ocorre 35 vezes (18 como verbo; 11 como substantivo e 6 como objeto e como complemento de preposição). Uma análise de suas colocações, no entanto, mostra que a preferência semântica é justamente a negação de **fear** e não seu reforço em 10 dos verbos e em 7 dos substantivos. Portanto, em 50% dos casos, **fear** é negado ou questionado (e.g. “Fear not”, “What need I fear?” “Hang those that talk of fear”, “nor shake with fear”). Knight também sustenta a relevância de **honour** (1968: 139). No entanto, **honour** somente ocorre 11 vezes na peça (cf. Tabela 3 acima).

Uma investigação da ocorrência de **fear** em *Macbeth* e no *corpus* de Holinshed pode contribuir também para a avaliação da qualidade dos dois textos. Holinshed é menos criativo. Há 29 ocorrências da palavra, mas, à diferença de *Macbeth*, a maioria se coloca como objeto de preposição (17). Holinshed usa o termos somente 4 vezes como verbo e 8 como substantivo. Seus colocados à esquerda são altamente previsíveis (“all”, “sudden”, “great”, “stand in”). Além disso, Shakesperare cria imagens originais ao personificar **fear**: “the initiate fear that wants hard use”; “pale-hearted fear”. Holinshed se mantém banal (“his owne feare fantasieth”; “a sudden terror and deadlie feare”).

#### **d. Análise de cima para baixo**

Ao invés de iniciar com uma lista de frequência, pode-se tomar um tema recorrente entre os críticos e examiná-lo à luz dos dados lingüísticos que o computador permite visualizar. Aqui, investiga-se, por exemplo, o papel de Lady Macbeth como mulher. A leitura da peça mostra que ela sempre invoca a masculinidade de Macbeth para questioná-la e assim provocar o marido. A palavra **man** ocorre em 34 casos. Uma concordância desse item lexical mostra que os colocados mais frequentes são **a** (14) e **of** (13), o que indica uma tendência à referência genérica. Em 10 casos, o verbo **be** se coloca com **man** e em 4 destes usa-se a forma comparativa (“like”, “as”, “more”). A análise da área semântica dos

colocados indica que em 12 instâncias, **man** está relacionado a poder e em 6, à temporalidade. Em Holinshed, a prosódia semântica positiva (Louw, 1993) de **man** é muito semelhante ao uso que Shakespeare faz, mas em muitos casos as comparações e associações à poder e temporalidade não são traçadas. Os epítetos somente ajudam a caracterizar a pessoa (“of great nobility”, “of loftie courage”, “well learned”, “well languaged”, “of great experience”). Na peça, o uso da palavra **man** aponta para a perspectiva renascentista de homem como centro de referência, como um paradigma, embora sujeito à temporalidade. Para levar o marido à ação, Lady Macbeth questiona justamente sua masculinidade. A palavra que ela mais utiliza é *great* (Louw 1991: 172) e geralmente com ironia. Esse achado contradiz a afirmação de Knight (1968: 141) de que Lady Macbeth ganha por apelar principalmente para o valor de Macbeth.

Por outro lado, enfatiza-se a capacidade reprodutora de **woman**, colocando-a em um contexto de prosódia semântica neutra. Esse uso é típico tanto em Holinshed quanto na peça. Há 13 ocorrências de **woman** em **MAC**. A Tabela 8 abaixo mostra a lista de seus colocados mais freqüentes. Direita e Esquerda se referem aos colocados posicionados com relação ao nóculo:

MAC		HOL	
Direita: born = 4	Esquerda: of = 8 born = 4 to = 3 a = 3	Direita: a = 10 of = 8	Esquerda: any = 5 borne = 3 poore = 3

Tabela 8: Colocados de “woman”

Esta tabela traz à luz o uso do adjetivo *poore* em **HOL**. Quando não são apresentadas em sua capacidade reprodutora, as mulheres são destituídas, empobrecidas de tudo, em contraste com o homem, com colocados à esquerda como *everie* (16), *noble* (9), *great* (8), *anie* (7). Os adjetivos *noble* e *great* contrastam com *poore*, da mesma forma que *any* contrasta com *everie*. De acordo com Quirk et al. (1980: 365), *any* pode expressar uma quantidade indefinida, enquanto que *every* é claramente singular. Portanto, aqui descreve-se o homem como único, com singular. Além disso, há duas ocorrências de

mulher em *Macbeth* associadas a partes do corpo: *woman’s breasts*; *woman’s ears*, o que não ocorre em Holinshed. A Tabela 9 abaixo compara a prosódia semântica de **woman** em *Macbeth* a três outras tragédias e à obra de Holin shed:

	Negativo	Positivo	Neutro	Total
<b>MAC</b>	4	0	9	13
<b>HAM</b>	5	0	3	8
<b>OTH</b>	5	7	4	16
<b>HOL</b>	4	5	8	17
<b>LR(Folio)*</b>	3	0	5	8

\*Aos invés da Edição Quarto, como nas tabelas anteriores, foi usada a

Edição Folio, que tinha uma instância a mais.

Tabela 9: Prosódia semântica de “woman”

Estaria, então, a prosódia semântica de **woman** em *Macbeth* indicando que, através dos atos de Lady Macbeth, Shakespeare condena mulheres que recusam seu papel de procriar? Seria essa recusa um crime hediondo que transformaria Lady Macbeth na 4ª bruxa? Nas palavras de Freud (1968: 135), “Acredito na *doença* de Lady Macbeth, na sua insensibilidade, através da qual ela está convencida de sua *impotência* contra os desígnios da natureza., e ao mesmo tempo em que é lembrada que é sua própria culpa se o crime a roubou da melhor parte de seus frutos” (meu itálico). Esse argumento é um tanto contraditório. Se ela é impotente contra os desígnios da natureza, se, como o *corpus* ilustra, ela é socialmente representada como um pobre elemento de procriação, por que não pode gerar filhos? Pelo contrário, ela consegue justamente aquilo pelo qual lutou. O que ela não pôde medir *a priori* foram as conseqüências de seu ato. Sua miopia não a permite antecipar os efeitos do crime no marido e nela mesma. Ela não percebe que o crime os separará, como de fato ocorreu. Eles passam a viver em mundos diferentes e, com isso, ela quebra emocionalmente. Seu suicídio resulta da solidão em que passa a viver. A análise de *corpus* permite uma visão diferente da personagem. Ao invés da esposa ambiciosa que Holinshed retrata, Lady Macbeth de Shakespeare é lingüisticamente representada como um personagem mais psicologicamente complexo: um dos primeiros exemplos da “louca no sótão” (Gilbert & Gubar, 1988). A complexidade da mulher retratada, seu relacionamento com o marido e a tensão dramática criada pelo uso que Shakespeare faz da linguagem, à diferença das soluções encontradas por Holinshed, ajudam a determinar a qualidade literária da peça.

## 7. Conclusão

Neste artigo mostrou-se como a lingüística de *corpus* pode contribuir para a interpretação de textos literários. Para tanto, tomou-se um texto canônico com uma longa história de literatura crítica para observação através de um software apropriado. Os critérios adotados foram a (im)previsibilidade e da função com o objetivo de verificar a qualidade literária dos textos. Cabe lembrar que atribuir qualidade literária é um processo complexo e que somente esses dois critérios não podem oferecer uma resposta final. Outros parâmetros podem ser usados para se entender porque Shakespeare se tornou canônico enquanto Holinshed, sua fonte, somente apresenta interesse histórico. Nesse caso, a análise de *corpus* deve ser enriquecida com outras abordagens, como estudos de narratologia e análise das estratégias de envolvimento, por exemplo, para verificar a tensão criada pela relutância dos dois protagonistas em levar o crime adiante, o papel mais proeminente de Lady Macbeth na peça, o acesso ao mundo privado dos protagonistas através de cartas trocadas e lidas em voz alta. Esses são outros elementos que contribuem para garantir a qualidade literária do tema em Shakespeare.

Como metodologia, a análise de *corpus* pode ser muito eficaz na sala-de-aula, promovendo independência de crítica. Ela permite que os alunos identifiquem, descrevam, generalizem, organizem, comparem, atribuam significado e, por fim, avaliem a qualidade da linguagem utilizada. Conforme comenta Miall (1991: 7-8), “a flexibilidade do computador como um instrumento, unido ao fato de que a atenção à tela desvia a atenção dada ao professor, encoraja os alunos a se engajarem mais ativamente no debate, central à área de Humanas”.

A análise oferecida não se pretende final. Cabe como um exercício preliminar a ser criticado e contextualizado. No entanto, algumas postulações podem ser feitas. Entre as várias vantagens, a análise de *corpus* permite aos estudiosos formular novas perguntas, que podem, inclusive, contradizer leituras tradicionais de um determinado texto e colocar novamente em julgamento sua qualidade literária. Essa metodologia é altamente verificável e proporciona acesso rápido a um número muito grande de textos. Com isso, facilita-se a análise intertextual. Na sala-de-aula, a análise de *corpus* propicia um ambiente onde o foco muda do professor para o aluno, que, por sua vez, passa a formular perguntas e a decidir o que deve ser investigado. Através de projetos em grupo e das discussões com seus pares, com o professor no papel de interlocutor, os alunos resistem a interpretações pré-formuladas e desenvolvem suas pesquisas de forma independente. Em suma, a análise de *corpus* permite a polifonia de discursos, um diálogo entre textos que pode, utilizado de forma apropriada, vir a ser um método transformador no ensino de literatura.

### Notas:

<sup>1</sup> Nota no Grupo de Discussão [BELMORE@vax2.concordia.ca](mailto:BELMORE@vax2.concordia.ca) em 6 de julho de 1998.

<sup>2</sup> Comunicação pessoal, junho de 1998.

<sup>3</sup> Devido à dificuldade de se conseguir a versão eletrônica completa de *Chronicles of England, Scotlande, and Irelande*, escaneou-se *The Words of Holinshed as read by Shakespeare*, editado por A. Nicoll & J. Nicoll, London: J.M. Dent & Sons Ltd., 1843. Trata-se de uma versão resumida do original, mas com uma fonte legível pela máquina. O processo resultou em um *corpus*

de 161.755 palavras. Nossos agradecimentos ao Programa de Pós-Graduação de Psicologia da UERJ, em especial ao Prof. Rodolfo Ribas, que contribui para a construção desse *corpus*.

<sup>4</sup> Agradecemos a Geoff Bambrook, que enviou essas listas iniciais por correio eletrônico e mais tarde ajudou na análise em uma visita à Universidade de Birmingham em julho de 1997, visita essa que contou com o apoio da Fundação de Amparo à Pesquisa do Rio de Janeiro (FAPERJ).

<sup>5</sup> i.e., a Edição Quarto. LRQ é o nome do arquivo na Edição Eletrônica da Oxford..

<sup>6</sup> “Um nóculo de uma colocação é a palavra cujo comportamento lexical está sendo observado no momento” (Sinclair, 1991: 175). Um espaço é “uma medida, em palavras, do co-texto de uma palavra selecionada para estudo. Um espaço de -4, +4 significa que 4 palavras ao lado de cada nóculo será considerada relevante para o contexto verbal” (Sinclair, 1991: 175). N-3 e N2

se referem à 3<sup>a</sup> e à 2<sup>a</sup> palavra à esquerda do nóculo, respectivamente.

<sup>7</sup> cf. II.iii.119

<sup>8</sup> Encontrados somente uma ocorrência em HCE e 8 em HOL (3 substantivos e 5 verbos).

**Bibliografia:**

AARTS, J. & Meijs, W. (1986) *Corpus Linguistics: Recent developments in the use of computer corpora in English language research*. Amsterdam: Rodopi.

BARNBROOK, G. (1996) *Language and Computers*. Edinburgh: Edinburgh University Press.

BRUNER, J.S.; Postman, L. & Rodrigues, J. (1951) “Expectation and the perception of color”. *American Journal of Psychology*, 64: 216-227.

CARTER, R.A. (1987) “Vocabulary and second/foreign language teaching”. *Language Teaching*, January, 3-16.

CARTER, R. A. (1997) *Investigating Literary Discourse*. London: Routledge.

CULLER, J. (1975) *Structural Poetics*. London: Routledge & Kegan Paul.

CULLER, J. (1983) *The Pursuit of Signs. Semiotics, Literature. Deconstruction*. London: Routledge & Kegan Paul.

ECO, U. (1992) *Interpretation and Overinterpretation*. S. Collini (ed.), Cambridge: Cambridge University Press.

FREUD, S. (1968) “Some character-types met with in psycho-analytical work”, in *Shakespeare. Macbeth. A Casebook*. J. Wain (ed.), London: Macmillan.

GILBERT, S. & Gubar, S. (1988) *The Madwoman in the Attic: The Woman writer and the nineteenth century literary imagination*. New Haven: Yale University Press.

GOMBRICH, E. H. (1986) *Art and Illusion: a study in the psychology of pictorial representation*. Washington, D.C.: The Trustees of the National Gallery of Art.

KNIGHT, W. G. (1968) “The Milk of Concord: Life-themes in *Macbeth*”, in *Shakespeare. Macbeth. A Casebook*. J. Wain (ed.), London: Macmillan.

KNIGHTS, L.C. (1933) *How many children had Lady Macbeth? An essay in the theory and practice of Shakespeare Criticism*. Cambridge: The Minority Press.

LOUW, B. (1991) “Classroom concordancing of delexical forms and the case for integrating language and literature”, in *Classroom Concordancing*. T. Johns & P. King (eds.), *ELR Journal*, 4, 151-178.

LOUW, B. (1993) “Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies”, in *Text and Technology*. M. Baker, G. Francis & E. Tognini-Bonelli (eds.), Amsterdam & Philadelphia: John Benjamins, pp.157-176.

LOUW, B. (1997) “The role of corpora in critical literary appreciation”, in *Teaching and Language Corpora*, A. Wichmann et alii. (eds.), London: Longman, pp. 240-251.

LURIA, A.R. (1976) *Cognitive Development: Its Cultural and Social Foundations*. Cambridge, Mass: Harvard University Press.

MIALL, D. S. (1991) “Introduction”, in *Humanities and the Computer*. D. Miall (ed.), Oxford: Oxford University Press.

- QUIRK, R. et al. (1980) *A Grammar of Contemporary English*. London: Longman.
- ROSENBLATT, L.M. (1983) *Literature as Exploration*, NY: MLA.
- SCHMIDT, S.J. (1989) "On the construction of fiction and the invention of facts". *Poetics 18*: 3 19-335.
- SINCLAIR, J. McH. (1982) "Lines about lines", in *Language and Literature*. R. Carter (ed.), London: George Allen & Unwin.
- SINCLAIR, J. McH. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SINCLAIR, J. McH. (1996) "The search for units of meaning", in *Textus IX*, no. 1, January-June, pp. 75-106
- SINCLAIR, J. McH. (1997) "Lexical Item". Unpublished monograph.
- SINCLAIR, J. McH. (ed.) (1987) *Looking Up*. London & Glasgow: Collins ELT
- van Peer, W. (1986) *Stylistics and Psychology: Investigations of Foregrounding*. London: Croom Helm .van Peer, W. (1997) "Toward a poetics of emotion", in *Emotion and the Arts*. M. Hjort & S. Laver (eds.), Oxford: Oxford University Press, pp. 215-224.
- WILLIS, D. (1990) *The Lexical Syllabus*. London & Glasgow: Collins ELT.

**Textos Eletrônicos:**

- The Helsinki Historical English Corpus (ICAME CD-Rom).
- Shakespeare, William, *The Complete Works*. Oxford Electronic Texts.
- Scott, M. & Johns, T. *Microconcord*. Oxford University Press, 1993.

*Sonia Zyngier é Ph.D. em Lingüística Aplicada pela Universidade de Birmingham e Mestre em Literatura Inglesa pela Universidade de Liverpool. Membro da Associação de Lingüística e Poética (PALA) e coordenadora do Projeto Integrado Internacional de Pesquisa REDES, vinculado ao Programa Interdisciplinar de Lingüística Aplicada da Universidade Federal do Rio de Janeiro, 'a Universidade de Munique e 'a Universidade de Kiev. Têm vários trabalhos publicados na área de estilística pedagógica, análise do discurso e conscientização literária.*