

Análise dos 13 contos de Carlos GASPAR

por André CAMLONG

Vamos tratar de analisar os 13 contos de Carlos GASPAR, utilizando o software STABLEX, cujo objetivo é proporcionar uma “análise lexical, textual e discursiva”.

Os 13 contos de Carlos GASPAR, escritor e jornalista maranhense, estão disponíveis na internet, constituem um corpus ideal, já que se trata de textos de dimensão razoável para serem tratados exhaustivamente, sem perder de vista o conteúdo narrativo, que pode ser facilmente memorizado.



O software STABLEX foi editado e comercializado pela Pirus de São Paulo: pedropaulo@pirus.com.br ou www.pirus.com.br. No mesmo CD, ao lado do software, figura além de um texto de “Ajuda” à Instalação, um livro de Lexicometria, em que o leitor vai encontrar essencialmente dois tipos de informação: uma informação relativa aos modelos matemáticos explorados e uma informação relativa à utilização da estatística.

Que nos seja aqui permitido dizer rapidamente, que a estatística que vamos praticar, é uma estatística paramétrica, objetiva, científica e indutiva. Quer dizer uma estatística de descrição e de ajuda à interpretação. Dito disso, o nosso desejo seria que o leitor esquecesse de uma vez a dicotomia entre “quantitativo” e “qualitativo”, um “barbarismo” amplamente difundido e sem nenhum fundamento, nem lógico, nem matemático, nem estatístico, nem lingüístico.

Aqui então, no primeiro momento, vamos apresentar os dados estatísticos do corpus inteiro, destacando a ACP, Análise em Componentes Principais, cuja finalidade é salientar a característica fundamental da distribuição das 13 variáveis (ou seja, dos 13 contos), e a seguir modalizar a AFD, Análise Fatorial Discriminante, para podermos aprofundar as estratégias discursivas dos 13 contos.

No segundo momento, vamos abordar a análise lexical, textual e discursiva dos 13 contos, deixando-nos guiar pelos dados estatísticos da ACP, que vai salientar a relação dinâmica que existe entre os 13 contos, e da AFD, que vai destacar tanto a estrutura preferencial como a estrutura privilegiada do léxico, da qual depende a estrutura dinâmica do texto e do discurso, que vamos tratar de aprender e compreender no próprio ambiente natural, até entrar na própria intimidade do autor quando elabora os contos.

I. A estatística

Os dados estatísticos resultam do levantamento e do tratamento exaustivo dos 13 contos feito com o software STABLEX. Infelizmente, não podemos dar aqui todos os dados destacados pelo tratamento dos textos. Portanto iremos diretamente aos dados fundamentais, as tabelas representativas e aos gráficos correspondentes para ter uma “visão” exata e adequada da estrutura do corpus e da relação que existe entre os contos.

Vale a pena sublinhar com força o papel que a estatística desempenha nesse tipo de análise descritiva, objetiva e indutiva, quer dizer “científica”, ou “matemática” no verdadeiro sentido da palavra. A estatística dá a conhecer para melhor reconhecer. Ela é útil enquanto instrumento de observação, de descrição, de medição, de avaliação, de identificação, de justificação, de interpretação e de verificação da estrutura fenomenal inerente ao corpus analisado. Ela não diz nada a respeito da essência do fenômeno observado, mas ela “faz ver” e “faz conhecer” para “fazer reconhecer”. Ela é um instrumento ideal utilizado para analisar e interpretar os dados estatísticos fornecidos pelos modelos matemáticos e estatísticos, e possibilitados pela potência de cálculo do “computador” (no sentido etimológico da palavra).

1.1 Dados estatísticos dos 13 Contos

No quadro seguinte figuram:

- na primeira linha, a classificação dos contos.
- na segunda linha, o léxico ou número de ocorrências.
- na terceira linha, o vocabulário ou número de palavras diferentes.
- na quarta linha, os hapax ou número de palavras de frequência 1 próprias a cada conto
- na quinta linha, o vocabulário de frequência 1
- na sexta linha, a percentagem do vocabulário de frequência 1
- na sétima linha, a percentagem de repetição

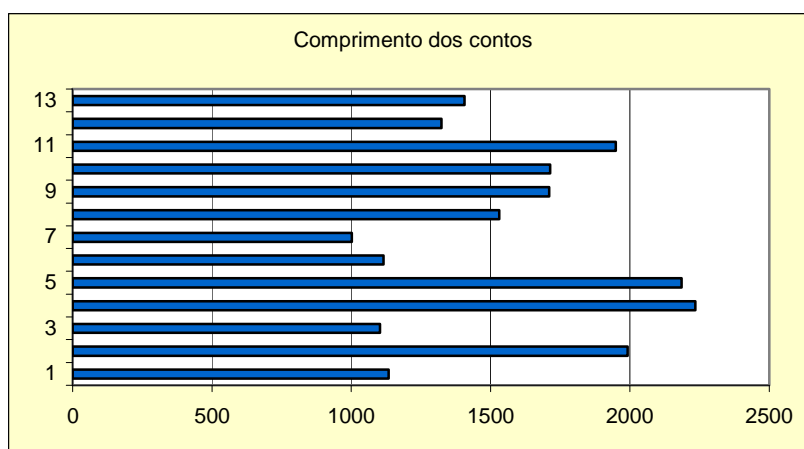
Contos	1	2	3	4	5	6	7	8	9	10	11	12	13	
Voc	20400	1134	1991	1103	2235	2185	1116	1002	1531	1710	1714	1949	1324	1406
Lex	5682	571	891	642	1061	1004	630	541	774	806	884	956	677	632
Hapax	3753	217	315	251	421	379	203	176	258	288	360	373	284	228
F=1		436	652	544	835	769	521	422	596	610	712	757	518	514
% de F=1		76,4	73,2	84,7	78,7	76,6	82,7	78	77	75,7	80,5	79,2	76,5	81,3
% de repetição		23,6	26,8	15,3	21,3	23,4	17,3	22	23	24,3	19,5	20,8	23,5	18,7

A nossa atenção é chamada pelo número elevado dos hapax de um lado, que representam mais ou menos 50% do vocabulário de frequência 1; e, de outro lado, pela alta percentagem do vocabulário de frequência 1, que vão sublinhando a importância do pitoresco na narrativa: o autor vai sempre à procura do detalhe, da minúcia, do pormenor.

Essa característica deverá ser observada de modo constante na análise dos textos e dos discursos.

Apesar de a estatística mostrar uma unidade na escritura e um parentesco evidente entre os contos, desde já podemos observar, através da riqueza do vocabulário, um texto acabado e uma narração elaborada prenunciando um discurso rigoroso e ponderado.

A representação gráfica da extensão dos textos dá uma idéia sintética e comparativa do comprimento dos contos:



Todos esses valores são diretamente tirados da primeira tabela de contingência, a TDF, a Tabela de Distribuição das Freqüências.

1.2 Densidades e peso lexical dos 13 Contos

A densidade representa o peso de cada conto dentro do corpus.

Os dados de densidade que figuram a seguir provêm todos da segunda tabela de contingência, a TDR, a Tabela de Desvios Reduzidos, automaticamente calculados pelo software numa planilha de cálculo, a partir dos cálculos de probabilidades efetuados na TDF.

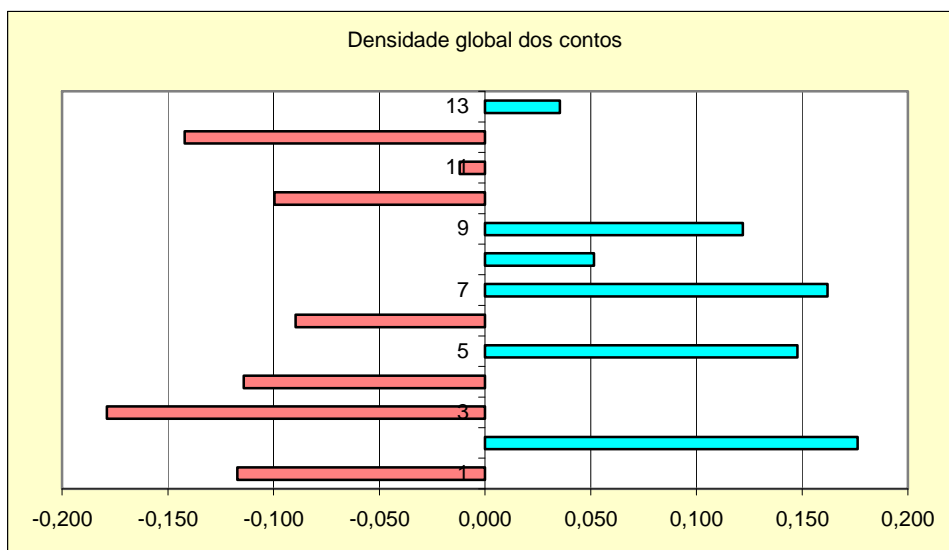
O leitor pode recorrer à definição dos programas inseridos no CD do software STABLEX para aprofundar as características dos cálculos estatísticos automaticamente realizados pelo computador, a partir do levantamento exaustivo do vocabulário dos 13 contos que constituem o corpus analisado.

É evidente que não podemos dar aqui o conteúdo integral das duas tabelas, mas podem ser solicitadas e conseguidas de boa graça.

Para simplificar a apresentação, damos a seguir o cabeçalho da dita TDR, para poder fixar os valores globais de densidade, assim como o valor do khi2 de Fisher, que define o grau de normalidade da distribuição e fixar os limites de controle da probabilidade.

Contos	1	2	3	4	5	6	7	8	9	10	11	12	13	
Total	-4,702	-9,490	14,273	-14,479	-9,236	11,968	-7,257	13,129	4,179	9,873	-8,058	-0,969	-11,501	2,866
Média	-0,058	-0,117	0,176	-0,179	-0,114	0,148	-0,090	0,162	0,052	0,122	-0,099	-0,012	-0,142	0,035
Khi2	0,195	0,014	0,031	0,032	0,013	0,022	0,008	0,026	0,003	0,015	0,010	0,000	0,020	0,001

O gráfico das densidades sob forma de barras permite discriminar automaticamente as densidades dos 13 contos:



À esquerda, cor de rosa, os contos cujo valor centrado reduzido é negativo, e, à direita, cor azul, os contos cujo valor centrado reduzido é positivo, conforme a escala que figura em baixo do gráfico.

Assim podemos ver que o conto número 1 tem um peso “negativo”, enquanto o conto número 13, o último, tem um peso “positivo”.

Quanto ao khi2 de Fisher, ele tem um valor altamente significativo de uma distribuição normal, perfeitamente centrada, já que o valor do khi2 = 0,195, com 13 graus de liberdade. De fato, com esse valor, a distribuição é perfeitamente controlada com uma probabilidade de 100%.

Qual é a significação desse khi2?

Esse grau de probabilidade significa que o corpus é um corpus de alta homogeneidade. Que todos os contos são da mesma veia, da mesma pena ou da mesma autoria. Aliás, quem pode confirmar essa avaliação a não ser o próprio autor Carlos GASPAR? Dizendo isso, deixamos ao leitor toda a liberdade de “avaliar” o peso da estatística na análise computacional.

De fato, no gráfico acima, vemos que os limites são apenas a $\pm 0,200$. Enquanto esses limites seria rejeitados a $\pm 3,565$ com uma probabilidade de 99,5%.

Daí resulta, pois, que o corpus dos 13 contos pode ser e vai ser analisado de modo preciso com o método estatístico, que vai permitir abrir os alicerces da análise lexical, textual e discursiva.

Infelizmente, aqui também não podemos abordar a observação da TDR, cuja grelha é um tesouro de “informação qualitativa” sobre o corpus, como vamos ver através da ACP (Análise em Componentes Principais) e da AFD (Análise Fatorial Discriminante).

II. A ACP ou a Análise em Componentes Principais

Que é a ACP? Ou, mais precisamente, que são os Componentes Principais?

Os Componentes Principais, do ponto de vista da estatística, são as variáveis consideradas como sendo uma unidade. No caso do corpus analisado, são os 13 contos, ou seja, as 13 variáveis que são representadas sob a forma de uma nuvem de 13 pontos.

Nessa primeira análise, vamos observar a distribuição dos 13 pontos e, sobretudo a representação das distâncias que separam os 13 pontos na configuração específica da “nuvem”.

Essas distâncias não são medidas de medição, mas medidas de posicionamento, de distância ou distanciamento entre os “pontos”.

É, a partir dos valores de correlação, que a Métrica R vai permitir medir, definir, descrever, identificar e observar as distâncias características do relacionamento entre os 13 contos.

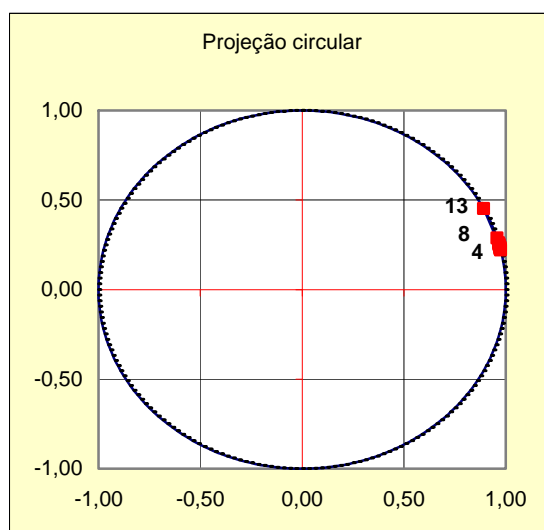
Quanto às características do modelo matemático da Métrica R o leitor vai encontrar a descrição no CD que acompanha o software STABLEX.

Dito isso, vamos aqui diretamente para os gráficos representativos da configuração espacial da nuvem de pontos.

Todos os valores são automaticamente calculados e todos os gráficos automaticamente realizados pelo software.

2.1 A projeção circular

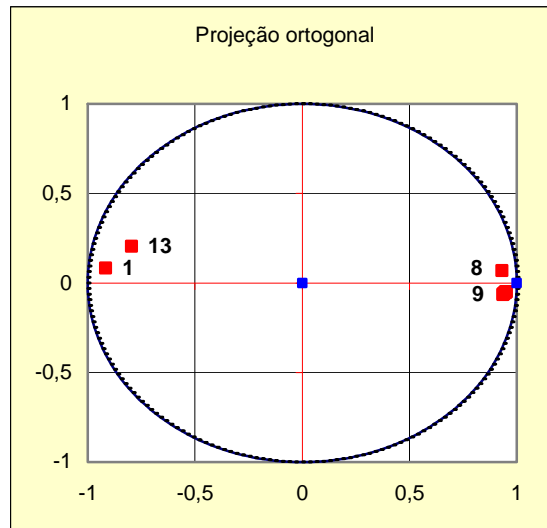
A projeção circular, a partir do co-seno da correlação, permite projetar a nuvem dos pontos no círculo trigonométrico:



A observação é imediata. Excepto o último conto, o número 13, os outros 12 contos são aglutinados, mostrando uma forte proximidade lexical. Assim destaca-se nitidamente a ligeira divergência do último conto, o conto da conclusão.

2.2 A projeção ortogonal

A projeção ortogonal, estabelecida a partir do cálculo dos quadrados do seno e do co-seno, em função da inércia total de distribuição, dá uma visão precisa da estrutura analítica da nuvem de pontos, já que o Alfa e o Omega, do primeiro e do último conto, ocupam, à esquerda, um lugar de destaque em oposição aos outros 11 contos. Estes, à direita do círculo, separam-se em dois grupos: o conto 8º com o seno positivo separa-se dos outros contos com um seno negativo.



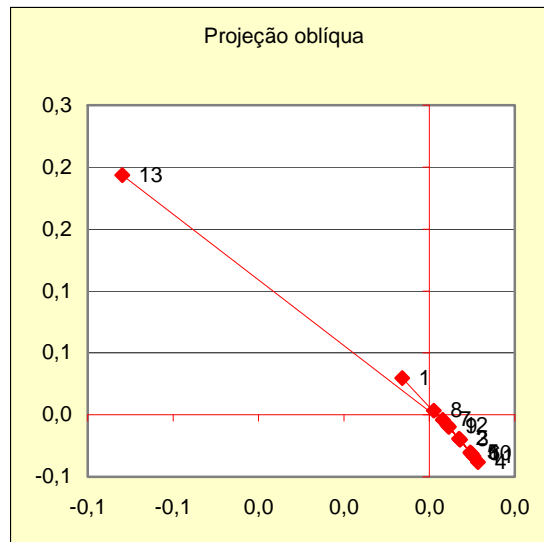
A projeção ortogonal dá assim uma visão mais precisa do posicionamento relativo dos contos dentro do corpus, designando desse modo uma ligação mais fina entre os léxicos, os textos e talvez os discursos.

Afinal são quatro grupos: dois à esquerda (os contos 1 e 13) e dois à direita (o conto 8º e os outros praticamente confundidos).

A ACP abre assim o caminho da AFD, que não pode deixar de lado essa observação fundamental. Esses resultados analíticos vão condicionar o processo de observação, de medição, de definição, de identificação próprio da análise lexical, textual e discursiva, e afinal garantir o fundamento da avaliação e da interpretação da essência do fenômeno observado.

2.3 A projeção oblíqua

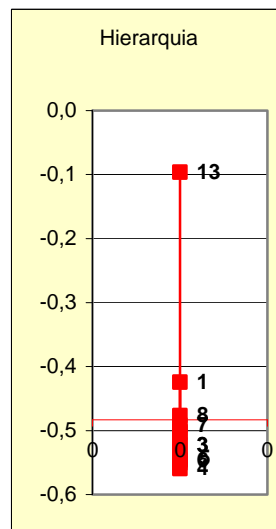
A projeção oblíqua dá ainda mais detalhes sobre as características fundamentais da distribuição representados sob um ponto de vista gráfico diferente.



Nesse caso, vemos nitidamente que o último conto ocupa um lugar de destaque altamente representativo duma diferença que ninguém podia supor à primeira vista, nem depois de uma leitura atenta e perfeitamente memorizada. O alfa e o ômega ocupam um lugar de destaque, cuja importância vai ser definida com a AFD.

2.4 A hierarquia

A hierarquia confirma a distribuição dos gráficos precedentes:



A distância do último conto, o conto número 13, é altamente significativa do valor conclusivo aí projetado pelo autor.

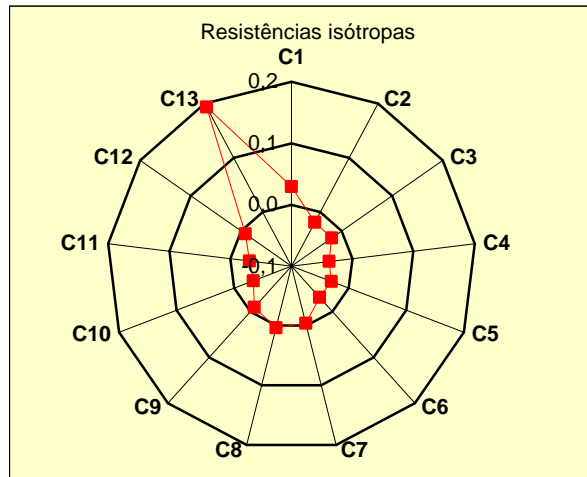
Em que consiste essa conclusão?

É preciso entrar na AFD para saber, conhecer e reconhecer de modo preciso.

Além disso, esse gráfico mostrar que afinal os contos são diferentes, já que formam uma seqüência abaixo da média (representada pela linha vermelha horizontal), indo do conto 8º praticamente na linha transversal até o conto 4º, na última posição.

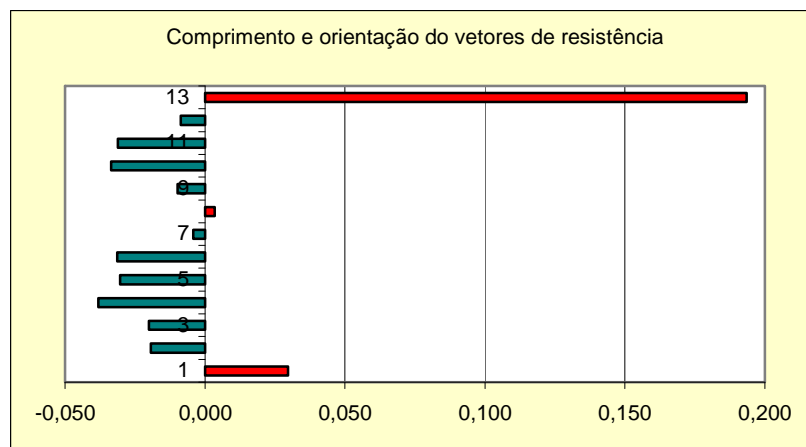
2.5 Os vetores isotropos de resistencia

Quanto aos vetores isotropos representativos da resistencia quando os valores trigonometricos sao reduzidos a inercia maxima ou a inercia total, eles mostram nitidamente a configuracao espacial dos textos, e, alem disso, os pontos de focalizacao do discurso.



Os contos C1, C8 e C13 projetam-se no exterior do círculo de raio 0, enquanto os outros contos projetam-se no interior do círculo de raio 0.

O gráfico sob forma de barras mostra com precisão o comprimento e a orientação dos vetores de resistência com a máxima inércia.



Para destacar ainda mais as diferenças, os vetores positivos de C1, C8 e C13 são de cor vermelha, e os vetores negativos dos outros contos são de cor verde.

Todos os valores destacados pela ACP têm um ponto de convergência que vai salientando a configuração estrutural do corpus analisado. O método objectivo e indutivo não vai formular nenhuma hipótese de trabalho, com faria o raciocínio dedutivo, raciocínio este totalmente

arbitrário, tautológico, “artificial e discutível”. Pelo contrário, o método indutivo abre todos os caminhos da “pesquisa”, sem preconceitos, sem prejuízos e sem compromissos.

Cada vez que vamos abordar a análise de um conto, já desde o início sabemos onde se situa no concerto das variáveis, qual é o peso, qual é a dimensão, qual é o posicionamento. A partir daquele momento pode-se entrar na análise fatorial discriminante.

2.6 Conclusão

De modo provisório, podemos salientar, a modo de conclusão, três pontos:

1. A estatística não pode ser reduzida ao “estatuto” (*status*, palavra etimológica de “estatística”) da população analisada.
2. A ACP é o momento primordial na descrição do corpus, já que mostra de modo evidente a relação que existe entre as variáveis: o distanciamento dos componentes.
3. Os modelos matemáticos explorados pela estatística abrem os alicerces do léxico, do texto e do discurso, garantindo a objetividade do percurso analítico da segunda etapa, o da AFD.

Podemos então abordar a análise lexical, textual e discursiva de cada conto, já que sabemos perfeitamente qual é o lugar que cada um deles ocupa no conjunto.

Para terminar essa conclusão, citaremos essa página de Nauro Machado que acompanha o livro dos contos, intitulado *A Invenção do Texto*:

Carlos Gaspar não busca no trabalho literário um sucedâneo descompromissado ou ameno para o seu duro e sério labor empresarial. As letras, no seu duplo sentido, lhe fornecem as possibilidades para reconciliar criativamente as pendências intercambiáveis da matéria e do espírito.

Este é um livro de contos, de crônicas, de reminiscências transfiguradas, ou, tudo isso é mais o resultado de uma invenção verbal necessária para a recriação do universo de quem o escreveu?

A começar pelo homem, que é "um invento de Deus", o autor, que acrescenta ser "um invento de meus pais", chega à afirmação conclusiva do conto número 13, com que finaliza no início o seu livro, corroborando o verso eliotiano de que no fim está o começo, na certeza também de que o por ele inventado está latente em si.

Este livro de Carlos Gaspar é a revelação daquilo que já intuía os leitores de seus artigos: o talento para transfigurar lembranças e fatos num contexto propício à sua desenvoltura ficcional.

É como se o narrado, ao fazer-se, cumprisse a estratégia não mistificadora a que ele se submete para uma melhor análise das situações, também psicológicas, dos seus personagens.

Não é, pois, matéria fácil classificar um livro como este, em que o fazer literário se realiza entre os limites daqueles diversos gêneros e que resulta por isso fora da compressiva uniformidade que se faz erroneamente complexa.

O livro se constitui em 12 relatos e o capítulo conclusivo (o conto 13 que principia o livro) de um autor que busca compreender os enredos em que ficcionalmente se envolveu, através de um fazer textual claro e objetivo, para chegar à realização contextual, que soa como metáfora do universo: esta invenção de todos.

A facilidade vocabular com que Carlos Gaspar desenvolve o enredo factual destes contos, na elegância de um estilo a denotar a convivência diária com os bons autores, e a montagem dessa

matéria, são elementos presentes neste livro, dando ao seu autor a plena posse afirmativa de um estilo pessoal.

Essa página levanta muitas perguntas, pelas quais podem ser formuladas várias hipóteses de trabalho que o método, nesse caso sim, permite verificar. Mas tal não é o objectivo fundamental do método indutivo e paramétrico: no primeiro momento, ele, graças à ACP, abre o leque dos componentes estruturais para posicionar todas as variáveis dentro do corpus, e, no segundo momento, graças à AFD, vai abordar a análise (*stricto sensu*) de cada componente.

Dado os resultados dessa primeira fase analítica, começaremos a análise pelo primeiro conto, o alfa, sabendo que o último conto, o conto número 13, é verdadeiramente o conto ómega, aquele que faz a síntese de um discurso que vai abrindo novas perspectivas, já que dá nova orientação à estrutura do corpus.

Assim esta análise vai seguir o seu próprio caminho sem preconceito nenhum, apenas passando pelos caminhos que acabam de ser revelados pela ACP.