

## Les études portugaises sur la langue parlée

Maria Fernanda Bacelar do Nascimento

Centro de Linguística da Universidade de Lisboa

Pour présenter un exposé à propos des études portugaises sur la langue parlée, je dois, forcément, les situer par rapport aux études sur corpus ; il y a une assez grande variété de types de corpus de portugais parlé dont, à ma connaissance, les trois les plus importants sont :

1. Les corpus dialectaux, c'est à dire, ceux qui s'insèrent dans la géolinguistique et ont été collectés pour permettre l'étude de la variation dialectale du point de vue phonétique, phonologique, morphosyntaxique et lexical, dans le territoire portugais. Leur objectif final est la réalisation d'atlas linguistiques nationaux ou régionaux où la variation dialectale est montrée de façon systématique<sup>1</sup>.
2. Les corpus de parole visant "l'entraînement et les tests de systèmes de synthèse et de reconnaissance de la parole ou, en général, la recherche fondamentale ayant pour but l'étude des relations entre les différents niveaux de représentation linguistique et les propriétés physiques des sons"<sup>2</sup>.
3. Les corpus oraux ou corpus de langue parlée, constitués par des ensembles d'enregistrements, soit d'entretiens spontanés soit de discours plus préparés ou plus formels et par leurs transcriptions orthographiques ; ils sont, en général, informatisés, de telle sorte que ces corpus peuvent être explorés et analysés simultanément ou en parallèle avec les corpus de langue écrite.

Nous avons, au Portugal, des corpus de portugais parlé, assez importants, de ces trois types et il y en a aussi au Brésil<sup>3</sup> et au Mozambique<sup>4</sup>.

---

<sup>1</sup> Segura da Cruz, M. L. 1996.

<sup>2</sup> Viana, M. C. *et alii*, 1996:189. C'est moi qui traduis.

<sup>3</sup> Castilho, A. T. *et alii*, 1996.

<sup>4</sup> Stroud, C. e Gonçalves, P., 1998.

Je limiterai mon exposé aux travaux réalisés sur le troisième type de corpus mentionné car c'est celui que je connais le mieux ; en effet, je coordonne une équipe qui travaille depuis longtemps à la construction d'un grand corpus général de langue parlée et de langue écrite, Le Corpus de Référence du Portugais Contemporain du Centre de Linguistique de l'Université de Lisbonne. Ce corpus a, maintenant, 92 millions de mots, dont près de 2 millions de langue parlée.

Il faut dire que, en général, les grands corpus de langue parlée transcrits orthographiquement sont assez rares ou, alors, assez réduits à cause des coûts qu'ils impliquent. Je crois, donc, que ce corpus oral du Centre de Linguistique de Lisbonne, de presque 2 millions de mots (que nous continuons à développer), constitue, compte tenu de sa dimension, et des critères rigoureux de transcription des textes, une base de données assez fiable pour les études sur le portugais parlé. Effectivement, plusieurs genres d'études ont été développés ayant recours à ces données et concernant les aires et les disciplines linguistiques les plus diverses, depuis les analyses lexicales pour l'établissement du Portugais Fondamental jusqu'aux analyses discursives réalisées par Maria Helena Araújo Carreira et aux analyses contrastives entre variétés du portugais. À propos des études sur le portugais parlé, il faut signaler le projet que je considère le plus important sur le portugais parlé réalisé au Brésil, sur le corpus NURC (Norme Urbaine Culte). Il s'agit de la Grammaire du Portugais Parlé, projet qui a été accompli sous la direction de Ataliba de Castilho. L'objectif de ce projet a été celui de préparer une grammaire référentielle du portugais parlé du Brésil, grâce à la description de ses niveaux phonologique, morphologique, syntaxique et textuel. Plusieurs volumes contenant des résultats de ce projet ont été publiés<sup>5</sup>.

Je reviens aux études que nous avons développées sur le corpus oral de Lisbonne. J'y distingue trois sortes de travaux :

- En 1970 nous avons commencé des études lexicales et quantitatives, visant l'établissement du vocabulaire du Portugais Fondamental, selon la méthodologie du Français Fondamental et de l'Espagnol Fondamental. Trois volumes ont été

---

<sup>5</sup> Voir dans la Bibliographie de cet article les volumes publiés dans le cadre du projet brésilien Gramática do Português Falado.

publiés : Le Vocabulaire, en 1984, et deux volumes de documents et de descriptions de la méthodologie suivie, parus en 1987<sup>6</sup>.

- Depuis 1987 nous avons réalisé des études sur quelques aspects de la syntaxe du portugais parlé. Ce sont des analyses en configurations selon les principes théoriques et méthodologiques proposés par Claire Blanche-Benveniste et l'équipe du Groupe Aixois de Recherches en Syntaxe. Je crois que ces études ont pu contribuer à dissiper les préjugés qui existaient à l'époque, chez nous, sur l'application de ce niveau d'analyse à la langue parlée. Dans le cadre de cette approche, nous avons surtout réalisé des études sur la syntaxe verbale,<sup>7</sup> sur les phénomènes de répétition lexicale et de structures syntaxiques<sup>8</sup>, des études sur les marqueurs discursifs<sup>9</sup>, et des études comparatives entre les textes de langue parlée et les brouillons de l'écrit.<sup>10</sup>
- Dans les dernières années nous avons développé au Centre de Linguistique des outils de traitement et d'exploration du Corpus de Référence qui visent, d'un côté, la disponibilisation des corpus et, de l'autre côté, l'extraction d'information lexico-grammaticale et statistique.

Nous jugeons utile de disponibiliser nos données linguistiques ainsi que nos outils d'exploration de ces données car les corpus sont envisagés de plus en plus par les linguistes comme un support solide pour l'application de méthodologies nouvelles plus efficaces ou, plus que cela, comme la substance même d'une nouvelle discipline linguistique qu'on appelle habituellement la Linguistique de Corpus.

Pour nous, il s'agit d'une nouvelle façon de faire la recherche linguistique, puisque les descriptions, les généralisations et les hypothèses théoriques que les études sur corpus électroniques permettent sont très solidement ancrées dans les données authentiques. Parfois cela remet en question quelques postulats de la linguistique conçus d'une façon intuitive.

---

<sup>6</sup> *Português Fundamental*, vol.I e vol.II, tomos 1 e 2.

<sup>7</sup> Bacelar do Nascimento, M. F. *et alii* 1987.

<sup>8</sup> Bacelar do Nascimento, M. F., 1996.

<sup>9</sup> Garcia Marques, M. L. , 1993.

<sup>10</sup> Raposo, H. , 1993.

À l'heure actuelle, le corpus de langue parlée est exploré et analysé simultanément avec le corpus de langue écrite dans le cadre de quelques programmes portugais et européens.

Faute de temps, je vais me limiter à la présentation de deux projets récents : l'un qui vise la disponibilisation de ces ressources linguistiques, l'autre qui vise l'extraction des associations de mots les plus remarquables pour l'analyse des occurrences lexicales et des expressions plus ou moins figées, ainsi que pour l'analyse des structures lexico-grammaticales récurrentes ou des tournures discursives qui sont propres aux différents genres textuels du portugais parlé et écrit.

En ce qui concerne la disponibilisation des données, je peux vous annoncer que très prochainement nous allons publier, avec l'appui de l'Institut Camões une partie des résultats d'un projet subventionné par le programme LINGUA/SOCRATES, coordonné par le Centre de Linguistique et ayant pour partenaires l'Université de Toulouse-le-Mirail (responsable : Paul Rivenc) et l'Université d'Aix-en-Provence (responsable Claire Blanche-Benveniste). De ce projet, sont en voie de parution deux sortes de publications : des CD-ROMs avec 83 échantillons de textes oraux et trois volumes d'études sur la langue parlée.

Nous croyons que la disponibilisation de ces textes oraux rendra service aux étudiants, aux enseignants et aux chercheurs qui travaillent sur le portugais parlé. Nous savons à quel point la notation orthographique de l'oral est inexacte et insuffisante, car, quand nous lisons les transcriptions des textes oraux il nous manquent toutes sortes d'indications indispensables à une bonne compréhension du texte, et, notamment, les marques prosodiques.

Les nouvelles technologies rendent possible à l'utilisateur des corpus oraux d'avoir sur un même support, et alignées, les productions orales et leurs correspondantes transcriptions orthographiques. Cela lui permet, tout d'abord, de faire sa propre écoute, car, souvent, il n'est pas d'accord avec la transcription proposée. Cette forme de présentation simultanée du son et de la transcription rend l'observation de la langue parlée plus vraie et plus naturelle. Au lieu de lire le texte oral (ce qui est toujours assez pénible et même assez bizarre), l'utilisateur l'écoute et, s'il le veut, il peut lire

en même temps sa transcription alignée. (Démonstration de 4 petits échantillons du portugais parlé au Portugal, à São Tomé, au Mozambique et au Brésil).

Au début ce projet était particulièrement dirigé aux étudiants de portugais langue étrangère, ayant pour but améliorer leur capacité de comprendre les variétés du portugais et aussi leur capacité d'expression. Maintenant on s'aperçoit que les CD-ROMs seront aussi très utiles aux professeurs et aux chercheurs.

Les CD-ROMs qui vont être publiés contiennent 83 enregistrements transcrits orthographiquement. Ces enregistrements, effectués depuis 1970 jusqu'à 1995, sont très diversifiés soit du point de vue des locuteurs, soit géographiquement (il y a des textes du Portugal – 30, du Brésil – 20, de l'Angola – 5, du Cap Vert – 5, de la Guinée-Bissau – 5, du Mozambique – 5, de S. Tomé e Príncipe – 5, de Macau – 5, de Timor-Leste – 3) soit, encore, du point de vue des thèmes et des types de rapports communicatifs. Cette diversité chronologique, géographique et communicative permet une perception dynamique de la langue et une compréhension de la remarquable unité structurale du portugais, malgré la diversité de ses réalisations nationales et régionales (diversité qui est essentiellement phonétique, prosodique et lexicale). Nous croyons que le développement de ce genre de ressources linguistiques mises à la disposition du public rendra possibles des études fondamentales sur les caractéristiques spécifiques de chacune de ces variétés du portugais et aussi des faits de langue qui leur sont communs et qui assurent l'unité de cette langue qui est parlée par presque 200 millions de personnes dans les 5 continents.

L'autre projet concerne les possibilités combinatoires des mots. Ce genre d'analyses est devenu classique dans le cadre des études linguistiques sur corpus. Il s'agit de faire, à partir d'un corpus donné, l'extraction automatique de suites de mots qui peuvent être, soit des suites "librement choisies"<sup>11</sup> soit des groupes de mots fortement figés, non-analysables du point de vue syntaxique et sémantiquement non-componentiels (traditionnellement només "expressions idiomatiques"), soit, encore, des mots qui cooccurrent assez systématiquement et dont les éléments constitutifs peuvent se trouver plus ou moins délexicalisés rendant évidentes les tendances d'un "principe idiomatique"<sup>12</sup> Des informations quantitatives et statistiques sur l'occurrence

---

<sup>11</sup> Sinclair, 1991.

<sup>12</sup> *idem.*

de ces associations lexicales rendent mesurables ces tendances “combinatoires ” jusqu’ici très peu étudiées au Portugal.

Le mot “combinatoire” implique une notion de régularité associative qui se manifeste par des faits quantitatifs, formels, sémantiques, distincts ou simultanés. Ainsi, nous avons des “combinatoires” caractérisées par leurs hautes fréquences mais qui ne reflètent aucun degré de déléxicalisation de leurs éléments constitutifs. C’est, par exemple, le cas de la co-occurrence “contar+história” (raconter+histoire) qui se produit souvent à l’oral et à l’écrit ou de la co-occurrence “dizer+assim” (dire+ainsi/comme ça) très fréquente à l’oral pour introduire les discours rapportés. (cf. Tableaux 1, 2, 3)

Dans ces deux exemples, les éléments de la combinatoire sont syntaxiquement libres ; il s’agit, dans les deux cas, de constructions verbales où les mots "história" et "assim" sont des éléments de réaction des verbes "contar" et "dizer". De même, chaque élément de la construction verbale conserve son sens indépendant, lequel contribue au sens global de l’expression. Par contre, d’autres co-occurrences lexicales présentent certains niveaux d’opacité sémantique et, corrélativement, des obstacles à leur décomposition en éléments syntaxiques. Selon leur degré d’opacité sémantique et d’analysabilité syntaxique, ainsi que de leurs fréquences et indices combinatoires dans les corpus, on peut induire qu’elles sont tout à fait figées ou qu’elles sont en voie de figement et donc plus ou moins proches d’un “principe idiomatique”. Ce sera, par exemple, le cas de la combinatoire “deitar+fora” qui a le sens global de “jeter” (cf. Tableau 4). L’analyse sémantique des éléments de cette combinatoire laisse voir que son sens global est moins prédictible que ceux des deux exemples précédents.

Les analyses de phénomènes d’associations lexicales restreintes ou élargies sur corpus permettent l’identification de modèles combinatoires assez significatifs du point de vue qualitatif et quantitatif, et montrent qu’il y a dans la langue beaucoup plus de “formules” toutes prêtes qu’on pouvait le croire. On peut conclure, aussi, que les frontières traditionnellement dressées entre faits de lexique et faits de syntaxe sont moins nettes que ce que les études fondées sur l’intuition présupposaient.

L'outil que nous avons développé pour ce genre d'analyses s'appelle "Dicionário de Combinatórias do Português" et il permet l'observation des comportements associatifs typiques des mots aux niveaux lexical, structural, sémantique et pragmatique.

L'utilisateur a accès à toutes les paires de mots de fréquence 4 ou supérieure à 4 dans un corpus donné et qui ont une distance entre ses éléments de 1 à 5 places à droite et de 1 à 5 places à gauche d'une forme-pôle. L'utilisateur a aussi accès à d'autres informations comme la fréquence des formes et des paires (qui sont lématés), aux indices combinatoires et aux concordances des paires de mots.

Ce sont, à l'heure actuelle, les explorations que nous sommes en train de faire au Centre de Linguistique sur le corpus de portugais parlé. Nous avons aussi commencé à faire des analyses contrastives sur les corpus oraux comparables de la variante européenne et de la variante brésilienne du portugais. Avec ces études sur le portugais nous espérons pouvoir contribuer aussi au développement des études comparées des langues romanes.

**Tableau 1 - "CONTAR+HISTÓRIA"** (CRPC: sub-corpus oral de Portugal)

JOGOS... DESENHAM COMO QUEREM, COMO QUEREM. ÀS VEZES ACIONAL, QUE É UMA **HISTÓRIA** DUMA VELHINHA, QUE O TIPO , E O TIPO COMEÇA A CANTAR, A FALAR, E NÃO SEI QUÊ, E LÁ ESTAVA ELA, COM UM OLHO FIXO EM MIM, NÃO SEI QUÊ, FALAR DA **HISTÓRIA** DE SANTA IRIA LEMBRAM-SE? QUE ATÉ É AS TANTAS **HISTÓRIAS** PARA CONTAR, NÃO SEI SE ESTÃO BEM NTANTO, AMBOS OS FILMES SÃO A MESMA COISA, AMBOS ELES E RESOLVEM FAZER UMA FESTA, OU DANÇAM OU CONVERSAM OU : TODOS OS DIAS À NOITE, ANTES DA MIÚDA ADORMECER NÓS CLIENTE APARECE, DIZ QUAL É O SEU PRODUTO - ISTO PARA UE TEM QUARENTA E TAL PÁGINAS, O (PROUST) DIZ QUE VAI ZER... QUER DIZER, NÃO É **HISTÓRIA** ASSIM QUE TENHA QUE ME MAIS PORQUE ERAM **HISTÓRIAS** QUE EU, QUE GOSTARIA DE SANTARÉM OU POUCO DEPOIS DE ESTAR EM SANTARÉM ELE VAI JÁ ESTAVA COMO A MAIOR PARTE DAS PESSOAS VELHINHAS A RO QUE MAIS TARDE, TEREI OUTRAS TANTAS **HISTÓRIAS** PARA MENTE POR USAR OS TERMOS DO POVO, NÃO É. EU POSSO-LHE A PARA O CHÃO, MAS DIZ QUE GOSTOU MUITO. NO (FUNCHAL) DOS PRIMEIROS CRISTÃOS. POIS. POIS, EU NÃO SEI BEM SE A BIOGRAFIA DE UM SENHOR QUE NÃO TENHA EXISTIDO E... TINUO, POR ENQUANTO AINDA NÃO LEVEI DENTADA" E DEPOIS ALA, ÉS TU OU SOU EU? BEM, FOI UMA **HISTÓRIA** QUE EU TE DA NÃO, MAS O FACTO É QUE A **HISTÓRIA** QUE HÁ BOCADO TE ADRID) OU DE (BARCELONA) , OU NÃO SEI QUÊ. DE MANEIRA QUÊ, PÁ, O (EGIPTO) E ENFIM TODAS AQUELAS COISA, PÁ, TIO... D: É FAMÍLIA DE UM LADO E FAMÍLIA DO OUTRO. C: UM SHOW... NÃO. EH, PÁ, QUE **HISTÓRIAS** QUE AQUELE TIPO RA QUE DÁ LOGO O DIREITO À GENTE ACHAR GRAÇA. E ENTÃO

**CONTA-SE-LHE** UMA **HISTÓRIA** E ATRAVÉS DESSA **HISTÓRIA** FAZER **CONTA**. E ENTÃO FAZ A VELHINHA, A VELHA FALA ASSIM: "AI O **CONTA HISTÓRIAS**, MAS **HISTÓRIAS** NO GOZO MAS COM O AR MAIS **CONTA** OUTRA VEZ UM BOCADO DA OUTRA **HISTÓRIA**, PERCEBES, OH **CONTADA**... NUM ATRAVÉS DUM... DUM ROMANCE DOS CHAMADOS RO **CONTADAS**, SE NÃO, DEVEM PECAR IMENSO POR EU AS TER RESUMI **CONTAM HISTÓRIA**, NÃO É, E AMBOS ELES QUE, PROCURAM PRODUZ **CONTAM HISTÓRIAS** E PASSAM ASSIM UMA TARDE OU UMA OU A NOI **CONTAMOS**, LÊ [...] , LÊMOS-LHE UMA **HISTÓRIA**. E ENTÃO CRI **CONTAR** A **HISTÓRIA** DO PRINCIPIO, NÃO É A: POIS! X: DIZ EXA **CONTAR** A **HISTÓRIA** DE UM AMOR DE (SWANN) , PORQUE EM CERTO **CONTAR** ASSIM MESMO A VALER, MAS QUER DIZER, HÁ SEMPRE UM **CONTAR**. ESTÁ CLARO QUE MAIS TARDE, TEREI OUTRAS TANTAS HI **CONTAR**... HA A **HISTÓRIA**, DE... DO NOME DE SANTARÉM... VAI **CONTAR HISTÓRIAS**, E MAIS PARA DIANTE JÁ AS REPETIA, DEPOI **CONTAR**, NÃO SEI SE ESTÃO BEM CONTADAS, SE NÃO, DEVEM PECA **CONTAR** UMA **HISTÓRIA** QUE SE PASSOU COM ELA QUE TEM MUITA P **CONTARAM-ME** A **HISTÓRIA** DELE TER INAUGURADO NO MESMO DIA D **CONTAREI** A **HISTÓRIA** QUE EU... PELO MENOS SEI... DE A TER **CONTE** A **HISTÓRIA** COMO QUISE" . "IMAGINE QUE HOUE UM DES **CONTEI-LHE** ESTA **HISTÓRIA**. ELA ACHOU IMENSA PIADA. ENTÃO.. **CONTEI** HÁ BOCADO E QUE VOU ENTÃO AGORA... FOI UM INDIVÍDU **CONTEI** RESUMIDAMENTE É ESTA. AH, ENTRETANTO ELE DEU O NÚM **CONTOU-NOS** A **HISTÓRIA** TODA. DIGO EU "BEM AQUI NÃO HÁ OUTR **CONTOU** ASSIM UMA **HISTÓRIA** E LEVOU CATORZE CONTOS POR AQUI **CONTOU** ESSA **HISTÓRIA**! X: NÃO FOI! JÁ UMA VEZ AQUI NESTE S **CONTOU**. ESTA FOI A QUE ME FICOU CÁ. A COISA QUE FUGIU... **CONTOU** UMA **HISTÓRIA** DA MULHER DE QUEM ELE GOSTOU E COM QU

Tableau 2 - "CONTAR+HISTÓRIA" (CRPC: Sub-corpus escrito de Portugal)

a, sei isso muito bem, mas essa **história** recuso-me a riando ao longo dos anos e até, num dos seus livros, sobretudo ligada ao sexo), na **história** que Arminda o lado, no coro desenvolvido dos atacantes... Henrique isas se passaram, porque, segundo o costume, cada um se a atitudes religiosas. A Grande Reportagem da SIC m ritmo de vida que deve ser lento e compassado. Não as? Eu logo vi. Estava a ler-tas na cara; pois agora de ouro. E depois, na venda do Fontinha, a **história** rece já perto e inadiável" (p. 161)), e a **história** ente, a manhã / tarde / noite, através das quais foi , reviver a cena, como nas **histórias** que as crianças s». Mas, o problema com os velhos é este: nunca nos meu rosto - disse Oriana. - A **história** que os homens izia o Sr. Prior no dia seguinte, em casa da Leonor, egressos, por vezes ajudava-me a adormecer, à noite, alquer transição, conheci-o há pouco no restaurante, is recente campanha do Omo máquina... E há mais. Vou ar a banda sonora: em vez dos discursos deles, tento a envenenava com ela. Ora imagine a Zefa da Graça a o-me a contá-la porque é uma **história** muito fácil de -lhe e conservou-o por algum tempo mudo. Esteve para , interromper o silêncio; ele imagina **histórias** para ogos Olímpicos de Atlanta já têm muita **história** para dotadas, mulheres. Estávamos cansadas. - Não podemos mprir com êxito a sua missão, deveria saber compor e arecia de casa, dias e dias seguidos; quando voltava avam na grande entrada vazia, onde os azulejos azuis rdavam, cosiam roupas no feno do alpendre, os homens ntara eu essas personagens junto de mim? Os desenhos Comité 4 de Fevereiro, ainda não havia Embaixada, e ar indiscreto de quem todos os dias se prostitui, e osaico envia-lhe daqui uns parabéns muito especiais. za que tem a cona apertada. Como tu gostas. Quem nos

**contá-la** porque é uma **história** muito fácil de contar. A Caroli **conta** a **História** de Portugal em verso. Apesar da facilidade co **conta** a João Carlos (onde sobressai a noção do olhar convencio **conta** a **história** idiota duma criança que vende almanaques, é a **conta** a **história** a seu modo: mas que a culpa é toda de Daniel, **conta**, esta noite, **histórias** de gente comum que meteu os pés a **conta** **histórias** farfalhudas, nem amargas, nada destoa, como ha **conta** tu o resto da **história** a essa gente, e que façam o favor **contada** a preceito e a derrota dos que tinham duvidado das var **contada** é, já, insuficiente, deixou " POR DIZER QUASE TUDO AFI **contada** uma **história**, perdem o sentido de espaço fechado, sonh **contam** a si próprias, sem fim, as cenas repetem-se, «devagarin **contam** as **histórias** que gostaríamos de ouvir, insistem em nos **contam** é verdadeira - disse a Mãe Natureza. - Mas falta-lhe me **contando** a **história** às velhas tias surdas, e acentuando a sua **contando** **histórias** sussurrantes; a fuligem elevava-se, formava **contando** uma **história** que não ouvi; «levas a Helena, fica em c **contar** a minha **história** no reino das audiências. Aconteceu há **contar** a minha **história**. - Mas quem é a senhora? - Eu sou a Mã **contar** a **história** e veja lá o que não diria! Clara ocultou o r **contar**. A Carolina do Ribeiro, mulher formidável, tenho a cert **contar** ao pai a **história** toda, mas calou-se. Tinha um coração **contar**, aqueles episódios bizarros e burlescos, como antigamen **contar** e... a procissão ainda não chegou ao fim. **contar** uma **história** sem palavras - disse uma. Com efeito, só a **contar** uma **história**, povoando-a de um certo tipo de personagem **contava** logo uma boa **história**, episódios aéreos com figuras hi **contavam**, com muitos detalhes realistas, **histórias** de idílicas **contavam** **histórias** de caça, de política, jogavam às cartas est **contavam** uma **história**, da esquerda para a direita. No entanto, **contei** a minha **história**, que naturalmente era uma **história** biz **conto-te** a sua **história**, sim, talvez inventada. Adormeces de n **"Conto** sempre a mesma **história** ... amo acima de tudo o rigor, **contou** esta **história** foi uma mulher medieval. Sua avó assistir

**Tableau 3 - "DIZER+ASSIM"** (CRPC: sub-corpus oral de Portugal)

E FIZ UM DE DOZE E DEPOIS R O CIGARRO PARA O CHÃO E DO FOI A ALTURA DO GRUPO, PERSPICAZ NESTAS COISAS, " E ELE OLHA PARA MIM E	<b>DIZ-ME ASSIM:</b>	"AI (MARIA) NÃO QUERO UM BOLO
	<b>DIZ-ME ASSIM:</b>	"MENINO, BAKA! BAKA!" QUER DIZ
	<b>DIZ-ME ASSIM:</b>	O FOTÓGRAFO: "MAS ENTÃO E AQUE
	<b>DIZ-ME ASSIM:</b>	"Ó MÃE, EU NÃO ACHO ISTO MAL,
	<b>DIZ-ME ASSIM:</b>	"PÕE-TE JÁ NA RUA". E EU DIGO
PARA A SECAR AQUI MELHOR. FILHOS TEM?» DIZ-ME ELA, TINHA MAIS ALGUMA COISA.	<b>DIZ-ME ELA ASSIM:</b>	"AH VOCÊ TRAZ FÉRIA MAS EU
	<b>DIZ-ME ELE ASSIM:</b>	«DOIS.» DOIS, EU SABIA QUE
	<b>DIZ-ME ELE ASSIM:</b>	"TENHO ALI UMA CÓMODA, GÉNER
QUE ELE TEM - E ENTROU E PERÍODO JÁ MALDISPOSTO, E	<b>DISSE-NOS ASSIM:</b>	"AH! ISSO COM OS CADERNOS,
	<b>DISSE-NOS ASSIM:</b>	"DE HOJE A OITO DIAS VOCÊS T
S ORDINÁRIO QUE O MEU. ELE E EU ESTAVA A CORTAR E ELA RRECIDO, E FOI PARA CASA E PRINCÍPIO - , CHAMOU-ME E UI POR DOIS MESES... E ELE OSTREI-LHE OS DOCUMENTOS E	<b>DISSE-ME ASSIM:</b>	"ENTÃO, MAS O QUE É ISTO? IST
	<b>DISSE-ME ASSIM:</b>	"JÁ CORTOU CABELO? JÁ TEVE A
	<b>DISSE-ME ASSIM:</b>	"MÃE, A... A VIZINHA JÁ LÁ NÃ
	<b>DISSE-ME ASSIM:</b>	"Ó (CARLOS), TU SABES O RESUL
	<b>DISSE-ME ASSIM:</b>	«Ó MÃE, DEIXE-SE ESTAR, DEIXE
	<b>DISSE-ME ASSIM:</b>	"OLHA, ESTE SENHOR É O SENHOR
ÃO CHEGAVA PARA COMER, EU MUITÍSSIMO BEM DISPOSTO, OMPRAR, VEJO TANTA GENTE, MAIS, ELE E MAIS QUATRO, O O GAJO COM MAIS QUATRO, U FICO CALADA, MAS DEPOIS TUA COLEGA (LEONILDE)". E É, NÃO SEI QUEM FOI E EU "PÕE-TE JÁ NA RUA". E EU DESPERTOU-ME CÁ AQUILO. E É ESSA?", É QUANDO EU LHE GNIFICADO DESSA PALAVRA!" ALAR COMO O MEU. ÀS VEZES ARA AMANHÃ SER O ENTERRO.	<b>DIGO ASSIM:</b>	«É MEIA HORA E ELE NÃO VEM»; TEL
	<b>DIGO ASSIM:</b>	«É UM FILME ESPANTOSO.» A: NÃO H
	<b>DIGO ASSIM:</b>	"EH, MEU DEUS, COMO ISTO ESTÁ MUD
	<b>DIGO ASSIM:</b>	"EH PÁ, ISTO HOJE VAI SER!..." IS
	<b>DIGO ASSIM:</b>	"EH PÁ, O QUE É QUE É ISSO?", "NÃ
	<b>DIGO ASSIM:</b>	«ELES TÊM RAZÃO, MAS EU COMO É
	<b>DIGO ASSIM:</b>	"ENTÃO DISSE-ME QUE ISTO ESTAVA..
	<b>DIGO ASSIM:</b>	"ENTÃO MAS, Ó OUVI LÁ PÁ, O QUE É
	<b>DIGO ASSIM:</b>	"ENTÃO QUE IDEIA AGORA É ESSA, NÃ
	<b>DIGO ASSIM:</b>	"MÃEZINHA, ACONTECE ISTO ASSIM, A
	<b>DIGO ASSIM:</b>	"MAS ENTÃO AINDA NÃO SABES O QUE
	<b>DIGO ASSIM:</b>	"MAS EU LÁ EM MINHA CASA OUVIA E
	<b>DIGO ASSIM:</b>	"MAS PORQUÊ?" SEI LÁ, OU MAIS DES
	<b>DIGO ASSIM:</b>	MAS QUEM É QUE FALA, ÉS TU OU SOU
I, EU QUANDO VI O PERIGO, ERES, FICAS EM CASA". MAS ELE: "A COMO É O POLVO?" ..), CALHAU?» «SOU, SOU.» LA.» «EU NÃO TENHO ISSO!» A UMA, UMA REMESSA DELES,	<b>DIGO EU ASSIM:</b>	«BEM... EH!» NAQUELE MOMENTO:
	<b>DIGO EU ASSIM:</b>	"BEM, PODE SER QUE LEVANTES AS
	<b>DIGO EU ASSIM:</b>	"É TANTO". MAS NÃO LIGUEI MAIS
	<b>DIGO EU ASSIM:</b>	«ENTÃO MOSTRE-ME LÁ A SUA CÉDU
	<b>DIGO EU ASSIM:</b>	«ENTÃO OIÇA LÁ E NÃO TEM UM
	<b>DIGO EU ASSIM:</b>	«ENTÃO QUAIS SÃO?» «OLHE MENIN
E É MAIS AS DUAS MOCINHAS. NTÃO QUANTOS SÃO, MENINA?»	<b>DIGO-LHE EU ASSIM:</b>	«ENTÃO E SÃO SÓ DOIS?» «ENT
	<b>DIGO-LHE EU ASSIM:</b>	«ENTÃO, VÁ LÁ DIZENDO OS NO
ALMA OLHO PARA A MENINA E ), QUANTO CUSTA ISTO?". E	<b>DIGO-LHE ASSIM:</b>	"E EU AINDA NÃO CONSEGUI PERC
	<b>DIGO-LHE ASSIM:</b>	"Ó, Ó..." - ESTAVA AQUI ELE,
AO FIM DO PERÍODO". E EU DA CABEÇA. EU OUTRO DIA,ÁRIO. QUEM ME DERA... AI, SCANSADO; CHEGOU CÁ ACIMA EU ENTÃO PAREI O CARRO E NORMAL. SÓ QUANDO ELE ME ISO TANTO DESTA EMPREGO".	<b>DISSE ASSIM:</b>	"AH SIM! ENTÃO TÁ BEM". CHATEOU-S
	<b>DISSE ASSIM:</b>	"AI ELE JÁ BEBE, AI QUE MAÇADOR".
	<b>DISSE ASSIM:</b>	"AI, QUEM ME DERA ANDAR SEMPRE A
	<b>DISSE ASSIM:</b>	«MÃE, O PAI ESTÁ NA FÁBRICA NÃO L
	<b>DISSE ASSIM:</b>	"MAS CONTE LÁ O QUE É QUE LHE ACO
	<b>DISSE ASSIM:</b>	"MAS O QUE É ISSO?! QUE CONVERSA
	<b>DISSE ASSIM:</b>	"MAS O SENHOR NÃO TENHA ILUSÕES,

NHOS QUE EU, QUE EU TINHA	<b>DISSE ASSIM:</b>	«OLHE, A MINHA MÃE TRAZ UM BEBÉ N
EGUINTE FUI TER COM ELE E	<b>DISSE ASSIM:</b>	"OLHE, PARA A OUTRA VEZ NÃO FAÇA
IAGEM ASSIM. A ENFERMEIRA	<b>DISSE ASSIM:</b>	«ONDE É QUE QUER IR ASSIM COM EST
OPA E DEPOIS O MEU HOMEM	<b>DIZ ASSIM:</b>	"NÃO, MULHER, REMEDIAMO-NOS COM..
A QUE ELE DESSE O VALOR.	<b>DIZ ASSIM:</b>	"O MEU PAI" - PORQUE CLARO QUE O
LÁ PEGA NA AGULHA. E ELA	<b>DIZ ASSIM:</b>	"Ó MÃE! EU QUERIA, MAS NÃO POSSO!
OMER À BICHARADA E O PAI	<b>DIZ ASSIM:</b>	«Ó RAPAZ, LEVANTA-TE!» ELE LEVANT
E ENTÃO O MEU MARIDO ATÉ	<b>DIZ ASSIM:</b>	«OLHA, FALAMOS JÁ TUDO ANTES DA F
.. TÊ-LOS CÃ, PARA...» E	<b>DIZ ASSIM:</b>	«OLHA, TAMBÉM DEIXOU CÃ A C OS DE
A E VOLTA. QUANDO VOLTA,	<b>DIZ ASSIM:</b>	"OLHE NÃO, NÃO ESTÁ". "TÁ A VER?"
QUE ERA BASTANTE GRANDE,	<b>DIZ ELE ASSIM:</b>	«AI, EU HEI-DE COMER AQUELE NO
SORTE, QUANDO EU O VI CÃ	<b>DIZ ELE ASSIM:</b>	"ENTÃO ESSE DEDO?", EU FUI VER
PONTUAL A CHEGAR A CASA.	<b>DIZ ELE ASSIM:</b>	«EU VENHO JÁ!» E TIROU O CASAC
PRONTO FUI BUSCAR OUTRA.	<b>DIZ ELE ASSIM:</b>	«NÃO TENHAS MEDO, NÃO ESTEJAS
ACOLÁ, ENCONTROU OUTRO E	<b>DIZ ELE ASSIM:</b>	"Ó PAI, Ó PAI, VENHA AQUI ASSI
EU SAÍ E QUANDO CHEGUEI,	<b>DIZ ELE ASSIM:</b>	"OLHA, AGORA JÁ ESTOU UM BOCAD
Á TEM HAVIDO PESSOAS QUE	<b>DIZEM ASSIM:</b>	"AH, ENTÃO, MAS ELE É UM FILHO
A FAZER ENGRAÇADAS E QUE	<b>DIZEM ASSIM:</b>	"AH, EU VIM DE LÁ TÃO ESTÚPIDA
UMA PALAVRA NOVA, DEPOIS	<b>DIZEM ASSIM:</b>	"AH!... ISSO É AGORA!", NÃO É?
MALA ÀS COSTAS. ÀS VEZES	<b>DIZEM ASSIM:</b>	"AI, EU NÃO GOSTO NADA DE VIAJ
POIS. X: MAS COMENTAM, E	<b>DIZEM ASSIM:</b>	«AI HOJE ESTÁ UM DIA LINDO!» C
ÃO, MAS ELAS VÊEM PORQUE	<b>DIZEM ASSIM:</b>	«AI QUE LINDO DIA!» C: O PROBL

#### Tableau 4 - "DEITAR+FORA" (CRPC: sub-corpus oral de Portugal)

<b>DEITOU</b> O INVÓLUCRO DO CIGARRO	<b>FORA,</b>	HOUVE UMA PESSOA QUE PEGOU, UM S
TAR ARRECADADA. É IR <b>DEITÁ-LOS</b>	<b>FORA,</b>	DAR-LHES DE COMER. DEPOIS VOLTO
CHEGOU AOS SETE MESES <b>DEITA-OS</b>	<b>FORA.</b>	AH, DIZEM OS MÉDICOS QUE É O ÚTE
UE SE QUER MAS DEPOIS <b>DEITA-SE</b>	<b>FORA</b>	A EMBALAGEM E ISTO TAMBÉM É MATER
FORA, É DESPERDIÇADA, <b>DEITA-SE</b>	<b>FORA,</b>	É UMA CAIXA DE MADEIRA QUE PÕE-S
TRO DIA, OUTRAS VEZES <b>DEITA-SE</b>	<b>FORA</b>	PORQUE NÃO SE VENDE... É ASSIM. N
UMA EMBALAGEM QUE DEPOIS <b>DEITA</b>	<b>FORA,</b>	É DESPERDIÇADA, DEITA-SE FORA, É
O, MAS AGORA FAZEM. <b>DEITAM-NOS</b>	<b>FORA.</b>	PARA COMPRAR SAPATOS. AH, ISTO H
POR ACHAR GRAÇA NUNCA <b>DEITÁMOS</b>	<b>FORA.</b>	NUNCA! ACHO QUE NÃO, ACHO QUE TÁ
QUE FICAR COM ELA PARA <b>DEITAR</b>	<b>FORA,</b>	MUITAS DAS VEZES. POIS, POIS, PO
NÃO TÊM MATERIAIS PARA <b>DEITAR</b>	<b>FORA,</b>	NÃO. SABE QUE SÓ, AQUILO SÓ COLA
LO QUE NÃO PRESTAVA <b>DEITAVA-SE</b>	<b>FORA.</b>	O... ESCOLHIDO À VISTA, ESCOLHID
ANDAM À SOLTA, AGORA <b>DEITEI-OS</b>	<b>FORA</b>	E ANDAM PARA AÍ, LOGO VÊEM, TORNA
TINHA-OS TODOS E DEPOIS <b>DEITOU</b>	<b>FORA</b>	E O (FRANCISQUINHO) FOI LÁ E TROU
E PARA RECORDAÇÃO NUNCA <b>DEITOU</b>	<b>FORA,</b>	O EXTINTOR NOVO ESTÁ ALI. NÃO! E
NOS PODEMOS <b>DEITAR</b> O DINHEIRO	<b>FORA</b>	DE QUALQUER MANEIRA, SEM SABER ON
EM, NENHUMA COISA, <b>DEITAR</b> NADA	<b>FORA</b>	- ATÉ PORQUE NÃO ESTAMOS EM TEMPO
TIVERAM QUE <b>DEITAR</b> A ÁGUA TODA	<b>FORA.</b>	REGARAM FLORES. "AFINAL A SENHOR

## BIBLIOGRAPHIE

ABAURRE, M. B. M. e RODRIGUES, Â. C. S. (orgs.) *Gramática do Português Falado*, vol. VIII. Campinas, Editora Unicamp, (No prelo).

ARAÚJO CARREIRA, M. H., *Modalisation linguistique en situation d'interlocution : proxémique verbale et modalités en portugais*, Leuvain-Paris, Peeters, 1997.

BACELAR DO NASCIMENTO, M. F., “Aspectos da Sintaxe do Português Falado (repetições lexicais e de estruturas sintáticas em produções orais : fenómenos de deslocação” *Actas do Congresso Internacional sobre o Português*, Vol. I, DUARTE, I. e LEIRIA, I. (orgs.) Lisboa, APL, Junho de 1996, pp 203-223.

BACELAR DO NASCIMENTO, M. F., *Contribuição para um Dicionário de Verbos do Português – novas perspectivas metodológicas*, Lisboa, CLUL/INESC, Dissertação para Progressão na Carreira de Investigação, (Inédito), 1997.

BACELAR DO NASCIMENTO, M. F., GARCIA MARQUES M. L. e SEGURA DA CRUZ, M. L., *Português Fundamental*, Vol.2 – *Métodos e Documentos*, Tomo I *Inquérito de Frequência*, Lisboa, INIC/CLUL, 1987.

BACELAR DO NASCIMENTO, M. F., RIVENC, P. e SEGURA DA CRUZ, M. L., *Português Fundamental*, Vol.2 - *Métodos e Documentos*, Tomo II *Inquérito de Disponibilidade*, Lisboa, INIC/CLUL, 1987.

BACELAR DO NASCIMENTO, M. F. RODRIGUES, M. C. e BETTENCOURT GONÇALVES, J. (orgs.) *Corpora*, *Actas do XI Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 1996.

BLANCHE-BENVENISTE, Cl. et JEANJEAN, C., *Le Français Parlé, Transcription et Edition*, Paris, INALF, 1987.

CASTILHO, A. T. de (Org.) *Gramática do Português Falado*, vol III, *As Abordagens*. Campinas, Editora da Unicamp/Fapesp, 1990.

CASTILHO, A. T. de (Org.), *Gramática do Português Falado*, vol. 1, *A Ordem*. Campinas, Editora da Unicamp/Fapesp, 1ªed. 1997.

CASTILHO, A. T. de e BASÍLIO M. (Orgs.) *Gramática do Português Falado*, vol. IV, *Estudos descritivos*. Campinas, Editora da Unicamp/Fapesp, 1996

CASTILHO, A. T., OLIVEIRA E SILVA, G. M. e LUCCHESI, D., “Informatização de Acervos da Língua Portuguesa” in BACELAR DO NASCIMENTO, M. F. *et alii* (Orgs.), 1996, pp. 113-128.

GARCIA MARQUES, M. L. (1993), *Os bordões no português falado contemporâneo. Análise de dois exemplos : Não é e Pá*, Lisboa, CLUL – Dissertação para Progressão na Carreira de Investigação, (Inédito), 1993.

HABERT, B., NAZARENKO, A. et SALEM, A., *Les Linguistiques de Corpus*, Paris, Colin, 1997.

ILARI, R. (Org.), *Gramática do Português Falado*, vol. II, *Níveis de Análise Linguística*. Campinas, Editora da Unicamp/Fapesp, 1992.

KATO, M. (Org.) *Gramática do Português Falado*, vol. V, *Convergências*, Campinas, Editora da Unicamp/Fapesp, 1996.

KOCH, I. G. V. (Org.) *Gramática do Português Falado*, vol VI. *Desenvolvimentos*, Campinas, Editora da Unicamp/Fapesp, 1996.

NEVES, M. H. M. (Org), *Gramática do Português Falado*, vol. VII. São Paulo-Campinas, Humanitas/Editora da Unicamp, 1999.

*Português Fundamental*, vol. I, Vocabulário e Gramática : Tomo 1 – Vocabulário, Lisboa, INIC/CLUL, 1984.

RAPOSO, H. P. *Enunciados orais e escritos em construção*, Lisboa, FLUL, Dissertação de Mestrado,(Inédito), 1993.

SEGURA DA CRUZ, M. L., “Os corpora dialectais” do CLUL : sua caracterização e objectivos”, in BACELAR DO NASCIMENTO, M. F. *et alii* (Orgs.), 1996, pp. 151-158.

SINCLAIR, J., *Corpus, Concordances ; Collocations*, Oxford, OUP., 1991.

STROUD C. e GONÇALVES P. (Orgs.), *Panorama do Português oral de Maputo* Vol. I – *Objectivos e Métodos*, Cadernos de Pesquisa nº 22 INDE, 1998, Vol. II - *A Construção de um Banco de Erros*, nº 24, INDE, 1998, Vol. III – *Estruturas Gramaticais do Português : Problemas e Exercícios*, nº 27, INDE, 1998.

VIANA, M. C. , TRANCOSO, I. MASCARENHAS, I., OLIVEIRA, L. C. e RIBEIRO, C. M., "Corpora de Fala em P.E., Constituição, Segmentação e Etiquetagem", in BACELAR DO NASCIMENTO, M. F. *et alii* (Orgs.), 1996, pp. 189-216.