

Maria Fernanda BACELAR DO NASCIMENTO

LE PROJET DU PORTUGAIS FONDAMENTAL

Le projet du Portugais Fondamental a été réalisé par un groupe de linguistes du Centre de Linguistique de l'Université de Lisbonne, d'abord sous la direction de M. L.F. Lindley Cintra.

Depuis 1978, M. Malaca Casteleiro a pris la direction de cette recherche. Dès le début, M. Paul Rivenc de l'Université de Toulouse a été le conseiller scientifique du groupe.

Ce projet s'est donné comme but une enquête statistique sur le portugais parlé en Europe, dont les résultats, sous le titre *Portugais fondamental, vocabulaire et grammaire* sont en voie de parution. Un vocabulaire fondamental de 2.217 mots a été constitué, ainsi qu'une grammaire de base fondée sur une sélection de faits de langue du discours oral relevés à l'intérieur du corpus de langue parlée et destinée à l'élaboration d'une grammaire pédagogique pour débutants.

Comme le Français Fondamental, le vocabulaire a été obtenu par le dépouillement de deux ensembles de matériaux :

- le Corpus de Fréquence (corpus d'énoncés parlés), de 700.000 occurrences recueillies dans un ensemble de 1.400 conversations enregistrées au magnétophone (qu'on appelle interviews), et

- le Corpus de Disponibilité (corpus de mots isolés, écrits), de 481.000 mots recueillis dans un ensemble d'enquêtes écrites par 620 témoins portant sur des centres d'intérêt spécifiques.

La liste de mots relevés dans ces deux corpus a subi l'analyse méthodique d'une commission internationale de spécialistes en linguistique et didactique du Portugais Langue Etrangère. Cette commission a décidé d'exclure un certain nombre de mots trop spécialisés et d'en ajouter d'autres considérés comme indispensables, soit pour remplir quelques lacunes observées dans la liste finale, soit pour compléter des séries du type : les jours de la semaine, les mois de l'année, les couleurs, etc.

L'utilisation de l'ordinateur nous a permis de donner aux enquêtes des dimensions beaucoup plus vastes que celles des enquêtes portées sur le français, l'allemand et l'espagnol. Par rapport au Français Fondamental, nous avons pu, d'ailleurs, introduire dans notre enquête des améliorations qui nous ont été conseillées par M. Paul Rivenc, d'après son expérience de co-auteur (avec M. Antonio Rojo-Sastre) de l'Espagnol Fondamental. Ces améliorations concernent surtout la constitution et les procédures d'analyse des deux corpus.

Le Portugais Fondamental contient un nombre beaucoup plus grand de mots que les autres "langues fondamentales", et il présente des "corpus" de dimensions différentes, comme on peut le vérifier en observant les tableaux suivants :

Tableau I

	Français	Allemand	Espagnol	Portugais
Nombre de mots du Corpus de Fréquence	312.135	595.157	800.000	700.000
Nombre de témoins	275	401	1.600	1.400
Nombre de mots du Corpus de Disponibilité	23.680	833.000	400.000	481.000
Nombre de témoins	704	5.300	800	620
Nombre de thèmes	16	21	25	30
Nombre total de vocables retenus	1.475	1.269	1.545	2.217

L'obtention d'un chiffre si grand (2.217 mots) s'explique surtout, non seulement par les vastes dimensions du corpus, mais aussi par le choix des seuils de fréquence, et par l'étendue des enquêtes de disponibilité (celles-ci, pour le portugais, portent aussi bien sur des substantifs que sur des adjectifs et des verbes). En outre, les additions de la commission du Portugais Fondamental, peut-être trop nombreuses quoique justifiées, ont aussi contribué à l'élargissement de nos résultats.

Tableau II

	Français		Allemand		Espagnol		Portugais	
Nombre total de vocables retenus	1.475		1.269		1.545		2.217	
Vocables fournis uniquement par les enquêtes de Fréquence	712	48 %	737	58 %	747	48 %	742	33,5 %
Vocables fournis uniquement par les enquêtes de Disponibilité	240 environ	16 %	348	27 %	518	33,5 %	645	29,09 %
Vocables fournis à la fois par les deux types d'enquêtes					202	13 %	405	18,26 %
Vocables ajoutés par la commission	520 environ	36 %	184	15 %	78	5,5 %	425	19,18 %

• LE VOCABULAIRE

Etant donné que ce stage a pour objectif l'étude du français parlé et des langues parlées en général, je me propose de décrire tout simplement, pour ce qui est du vocabulaire, notre corpus de langue parlée, en ce qui concerne les points suivants :

- Conditions d'enregistrement des interviews
- Leur transcription
- Méthodes suivies pour le découpage du texte
- Analyse des unités lexicales
- Indexation du vocabulaire

- Conditions d'enregistrement des interviews

Nous avons recueilli par enregistrement direct tous les textes utilisés dans nos dépouillements. On a enregistré 1.800 conversations dont nous avons sélectionné 1.400.

En général, les enregistrements ont été effectués au su des sujets, dans une ambiance qui leur était familière, ce qui permettait une détente favorable à l'expression spontanée. La conversation portait généralement sur des sujets de la vie quotidienne (famille, profession, problèmes économiques, etc.) qui leur tenaient à coeur.

On a interviewé 50 % d'hommes et 50 % de femmes d'âges compris entre 15 et 60 ans (en donnant plus d'importance à une tranche d'âge située entre 20 et 30 ans). Les sujets étaient autochtones de la région enquêtée ou alors ils y habitaient depuis 10 ans au minimum. Les témoins appartenaient à plusieurs niveaux de culture (depuis ceux qui ne savaient ni lire ni écrire jusqu'à ceux qui avaient une formation supérieure). Cependant les niveaux moyens ont été privilégiés.

L'enquête a été menée sur tout le pays (le continent et les îles) et le nombre de témoins par département a été fixé d'après les indices de densité populationnelle, en tenant compte de critères d'ordre socio-économique et culturel. On a surtout privilégié les centres urbains.

Tableau III : Enquête de fréquence

<i>Départements</i>	<i>Nombre de témoins</i>
Lisbõa	400
Coimbra	150
Porto	150
Viana do Castelo	40
Braga	40
Vila-Real	35
Bragança	35
Aveiro	45
Viseu	40
Guarda	35
Castelo-Branco	45
Leiria	45
Santarém	45
Portalegre	35
Setúbal	65
Evora	40
Beja	40
Faro	40
Açores	40
Madeira	35
Total	1400

- Système de transcription des enregistrements

Nous avons recueilli des textes prononcés par un seul locuteur ou par plusieurs locuteurs parlant entre eux. Nous avons noté les prises de parole des locuteurs et de l'enquêteur, mais, puisque notre premier objectif était l'analyse quantitative du lexique, et que les caractères statistiques de chaque interview devaient être respectés, nous avons sélectionné un seul locuteur par échantillon et un seul échantillon de 500 mots graphiques par locuteur. Il faut cependant souligner que nos interviews sont beaucoup plus étendues que cela, ce qui nous a permis de choisir les morceaux les plus intéressants du point de vue du discours spontané et de la diversification lexicale. La transcription des morceaux choisis a toujours été faite par l'enquêteur qui interviewait et cela dans le délai le plus court par rapport à la date d'interview.

Etant donné que nous n'avions pas l'intention de faire des études phonétiques ou phonologiques, nous avons choisi la transcription orthographique qui nous a semblé suffisante pour l'exploitation lexicale du corpus.

Dans la reproduction écrite des textes, on a essayé d'être aussi fidèles que possible à la production orale. Nous avons donc tout transcrit : les hésitations, les mots tronqués, les corrections, les répétitions et aussi les fautes apparentes même. Mais, comme vous le savez, la perception peu nette d'éléments prononcés très rapidement, ou très amalgamés à l'environnement phonique pose beaucoup de problèmes et a fortement contribué à la lenteur du travail de transcription. Nous avons décidé de transcrire uniquement ce qui était d'une perception auditive nette et de remplacer par un signe conventionnel (les points de suspension entre parenthèses) les unités incomprises. Nous avons évité de construire notre propre interprétation et nous n'avons pas proposé deux ou plusieurs solutions pour toutes les divergences auditives, comme vous le faites à Aix, car, comme je l'ai déjà dit, il s'agissait d'obtenir un corpus de lexique et non d'obtenir un corpus pour l'exploitation syntaxique. Nous avons cependant reconstitué les mots tronqués en mettant entre crochets la partie reconstituée du mot dans le cas où la reconstitution n'offrait aucun doute.

Ex. : "Eu não po[s so], não posso dizer".

En ce qui concerne les phatiques du type : *hã, hem, hum, oh*, nous les avons orthographiés de façon conventionnelle; les numéraux et les lettres qui représentent des quantités comme dans "n fois" ont été écrits en toutes lettres.

Nous n'avons pas suivi les règles de ponctuation de l'écrit mais on a essayé de représenter le rythme du discours du locuteur avec l'emploi des signes conventionnels de ponctuation, sauf si ce procédé conduisait à des ambiguïtés dans la compréhension du texte. On a aussi représenté les pauses très longues par des points de suspension et mis entre virgules les répétitions, les corrections et les expressions

(du type : "nãõ é, quer dizer", "n'est-ce pas", "tu vois"). Les propositions incisives ont été mises entre deux tirets.

Les réalisations phonétiques très déviantes ainsi que les rires, les cris, les interruptions incompréhensibles ont été notés ou expliqués en bas de page.

• L'édition du texte

Chaque texte contient, à part le nom de l'enquêteur, des renseignements d'ordre socio-culturel sur le témoin, en vue d'une ultérieure exploitation sociolinguistique de notre corpus. Ainsi, on a constitué une fiche qui décrit le nom et l'âge du témoin, la région où il habitait, son sexe, son métier, son niveau de culture, et finalement le nombre d'habitants du lieu de résidence (moins de 1.000 habitants, de 1.000 à 5.000, de 5.000 à 10.000; etc.) et des influences linguistiques subies, s'il y avait lieu.

Ensuite, sur la version orthographiée, on a préparé le texte en vue d'une analyse statistique du vocabulaire. C'est pour des raisons d'ordre statistique et aussi pour rendre valable la notion de *répartition* que nous avons établi des textes de longueur égale appartenant toujours à des locuteurs différents.

Voici les principales décisions qu'on a prises pour l'édition, concernant le but de l'enquête :

- barrer toutes les interventions qui n'avaient pas été produites par le témoin choisi;
- mettre entre parenthèses les noms de personne, les noms de lieu et aussi les citations, les proverbes, etc. qu'on ne comptait pas pour l'obtention des 500 mots de chaque texte;
- éliminer les mots ou les phrases répétés par hésitation
 Ex. : O armário *da, da, da* cozinha → O armário da cozinha.
 Si dans la répétition d'une phrase par hésitation le témoin avait ajouté un élément quelconque on a tout conservé.
 Ex. : Um amigo, portanto um amigo ...
- choisir, comme unité de lexique *le mot graphique*, quoique ce soit un critère difficile à suivre car la définition de mot reste à établir.

Nous avons donc considéré comme une seule unité, par exemple, les formes contractées (ex. : em + a → na). Par contre, nous avons considéré comme une seule unité les unités complexes qui ont subi le processus de lexicalisation, c'est-à-dire dont les constituants ont été progressivement soudés par l'usage et qui constituent déjà une suite indissociable exprimant une notion unique, indépendamment de la graphie d'usage :

Ex. : Fim-de-semana
Guarda-chuva
Mulher-a-dias

- choisir comme unité de texte, 500 mots graphiques, parce que c'est un nombre de mots qui correspond, dans la chaîne parlée, à un échantillon représentatif du récit du locuteur et, en même temps, parce que c'est un chiffre rond, ce qui est utile du point de vue statistique, permettant des calculs faciles et rapides.
- Edition du texte pour l'ordinateur

Nous avons donné un numéro d'ordre à chaque interview et nous avons codifié toutes les informations, déjà mentionnées, sur le témoin afin de les rendre éventuellement utilisables dans un certain nombre de dépouillements sociolinguistiques par ordinateur. Les textes, préparés pour l'ordinateur, qui ne dépassent jamais 500 mots, contiennent cependant tout ce qui est entre parenthèses pour rendre intelligibles les contextes, que, par la suite, on a demandé à l'ordinateur, en vue d'analyses grammaticales postérieures.

Nous avons élaboré un système de codes pour marquer les accents, les diacritiques et les signes de ponctuation.

Tableau IV

CODES	SIGNES
1 _____	accent aigu
2 _____ /	accent circonflexe
3 _____	accent grave
4 _____	signe de nasalisation
5 _____	cédille
6 _____	accent aigu + cédille
7 _____	accent circonflexe + cédille
8 _____	accent grave + cédille
9 _____	signe de nasalisation + cédille
\$ _____	point d'interrogation
% _____	point d'exclamation
/.../ _____	le début et la fin du discours direct
+...+ _____	le début et la fin des phrases incises
* _____	points de suspension

- L'analyse du vocabulaire

La mémorisation du corpus de 700.000 mots (500 mots multipliés par 1.400 interviews) a été réalisée d'abord sur cartes perforées, puis enregistrée définitivement sur bande magnétique, après quoi la machine nous a fourni une première liste imprimée automatiquement, contenant 25.107 formes différentes par ordre alphabétique et suivies du nombre d'occurrences attestées dans le corpus (cf. Annexe I).

Sur ces listes nous avons fait une lématisation provisoire des vocables au-dessus de la fréquence 40, en tenant compte, pour établir sa fréquence, du nombre d'occurrences de toutes ses formes flexionnées ou contractées, aussi bien que des formes composées et des augmentatifs, diminutifs, comparatifs et superlatifs des adjectifs.

Puisque nous n'avions que les données numériques brutes des occurrences de chaque forme, cette lématisation, dite provisoire, a été plutôt théorique. Pour établir les paradigmes, nous avons donc conçu les différentes hypothèses de réalisation grammaticale des homographes et ensuite nous avons demandé à l'ordinateur de nous fournir les

contextes de 65.000 formes susceptibles de réalisations homographiques (cf. Annexe II).

L'analyse d'homographes a été réalisée sur ces listages qui contenaient, d'ailleurs, des contextes assez longs pour nous permettre la classification grammaticale des formes lexicales. Pour ce faire, on a utilisé une terminologie tout-à-fait traditionnelle qui n'est pas incompatible avec les diverses écoles linguistiques de nos éventuels utilisateurs.

Sur les listes de contextes, après chaque forme analysée, on a mis un code qui désignait la classe grammaticale, ce qui a permis à l'ordinateur de classer ces formes séparément et de nous en fournir la fréquence.

D'autre part, nous avons fait le dépouillement des locutions prépositives, conjonctionnelles, pronominales et adverbiales du corpus de fréquence.

Pour isoler ces groupes, on a suivi les critères suivants :

- indissociabilité des éléments (sauf en ce qui concerne l'insertion entre ces éléments de quelques adverbes caractérisés par leur grande mobilité dans la phrase);
- valeur mono sémique du groupe;
- équivalence morphosyntaxique entre un groupe donné et la classe grammaticale correspondante.

L'analyse de ces locutions a été réalisée sur des listes de contextes (cf. Annexe III).

Après toutes ces analyses, on a pu faire la lématization définitive des formes, en éliminant toutes les possibilités théoriques non réalisées et en ajoutant les locutions grammaticales de fréquence égale ou supérieure à 40, qui correspond au seuil de fréquence établi pour le corpus de langue parlée.

On a demandé à l'ordinateur *la répartition* du paradigme des vocables de fréquence entre 60 et 40, ce qui nous a permis de décider avec plus de rigueur de son inclusion ou exclusion de la liste finale.

Tableau V

Lémathisation provisoire	Lémathisation définitive	Liste finale
ALTA (sub. fém.) - 59	ALTA (sub. fém.) - 2	
alta 44	2	
altas 15	∅	
ALTO (sub. mas.) - 70	ALTO (sub. mas.) - 13	
alto 60	11	
altos 10	2	
ALTO (adj.) - 131	ALTO (adj.) - 111	ALTO (adj.) - 111
alta 44	42	42
altas 15	15	15
alto 60	44	44
altos 10	8	8
altísimas	1	1
altíssimo 1	1	1
ALTO (adv.) - 61	ALTO (adv.) - 5	
alto 60	5	
altíssimo 1	∅	

• L'indexation des vocables

Nous avons indexé les formes choisies pour représenter le paradigme par ordre alphabétique, d'une façon très proche des entrées des dictionnaires : les verbes à l'infinitif, les noms animés au masculin singulier (au cas où ils ont un radical identique au masculin et au féminin), les adjectifs au masculin singulier, etc.

La catégorie grammaticale des vocables susceptibles d'homographie a été indiquée en face de chaque forme.

Nous avons donné quelques renseignements qu'on a jugés utiles sur les niveaux de langue, notamment l'usage de vocables familiers ou argotiques.

Quand il s'agissait de groupes de mots, le groupe apparaît indexé sous chacun de ses éléments (cf. Annexe IV).

• LA GRAMMAIRE DE BASE

Nous considérons que toute grammaire pédagogique pour débutants doit être une grammaire du discours et c'est pourquoi nous avons relevé des faits de langue d'une régularité assez stable à l'intérieur d'un sous-corpus d'interviews de la région de Lisbonne.

Evidemment, on n'a pas eu l'intention d'élaborer un ouvrage didactique mais de faire *la description* d'un certain nombre de faits morphosyntaxiques de la langue parlée, lesquels constituent autant de données pratiques qui, à notre avis, seront très utiles à ceux qui travaillent sur les structures syntaxiques et les données d'ordre morphologique qui doivent être introduites dans une grammaire d'apprentissage.

Nous avons essayé de faire cette analyse directement sur les textes des interviews (dans certains cas nous avons utilisé les listes de contextes) afin de tenir compte, quand cela était possible, des conditions de la communication et des situations de l'énonciation associées aux énoncés.

Finalement, nous avons organisé ce chapitre par classes de mots, telles que : les noms, les adjectifs, les verbes (modes, temps, auxiliaires de temps, de mode, d'aspect); et par certains types de phrases, tels que : les phrases passives, les phrases interrogatives, les phrases négatives.

Le groupe chargé de l'élaboration du Portugais Fondamental a décidé de faire sortir tout de suite après la publication du Vocabulaire et de la Grammaire, une autre publication qu'on appellera *Portugais Fondamental, Méthodes et Documents*. Ce volume aura la forme d'un rapport très détaillé sur la méthodologie suivie dans notre première publication. D'autre part, nous voulons faire connaître un nombre assez important de matériaux concernant le portugais parlé contemporain.

En fait, l'élaboration du vocabulaire fondamental et de la grammaire de base nous ont permis de recueillir des corpus qui constituent une mine inépuisable pour la recherche linguistique, soit appliquée, soit théorique. Comme vous avez eu l'occasion de le voir, nous possédons des données de langue parlée très importantes qui peuvent donner lieu à la réalisation de travaux scientifiques, sur le portugais d'une part et d'autre part à l'élaboration d'études de type contrastif, entre langues parlées.

En ce moment, au Centre de Linguistique de Lisbonne, nous nous intéressons tout particulièrement aux méthodes d'analyse descriptive portant sur la langue parlée, en ce qui concerne surtout les relations entre lexique et syntaxe, du fait que nous avons l'intention d'élaborer un dictionnaire contextuel du Portugais Fondamental (dont la première partie sera consacrée aux verbes) en profitant des matériaux recueillis jusqu'à ce jour.

Maria Fernanda BACELAR DO NASCIMENTO

★

Annexe I

1	CARVALHO
25	CARVAO4
1	CARVAOZINHO4
1550	CASA
3	CASACA
1	CASACAO4
53	CASACO
14	CASACOS
31	CASADA
8	CASADAS
10	CASADEBANHO
25	CASADO
9	CASADOS
14	CASAIS
36	CASAL
1	CASALINHO
1	CASALITO
11	CASAM
2	CASAMENTEIRO
62	CASAMENTO
6	CASAMENTOS
7	CASAMOS1
1	CASANDO
1	CASAO4
46	CASAR
12	CASARAM
2	CASAREM
1	CASARIO
3	CASARMOS
1	CASAROES4
248	CASAS
4	CASASDEBANHO
1	CASASMUSEU
7	CASASSE
2	CASAVAM
19	CASCA
4	CASCAS
1	CASCATA
1	CASCO
2	CASCOS
1	CASE
27	CASEI
6	CASEIRA
1	CASEIRAMENTE
12	CASEIRO
4	CASEIROS
1	CASES
17	CASINHA

Annexe II : Extrait de la liste des contextes pour l'analyse des homographes

VOUZ DA VEP OJ ALISSI MSSI, LIVREINHOS, TAMBEM, E NOS OS LIVREINHOS, FU LIZO EM NIN PESSOALMENTE, TAMPENI MO PERTEI O U NINPARAS AS
 CONTAS, O 13 VOCABULO 243 ENTREVISTA 243 VAGA [04]
 NOS LINDS DA VERON E APRESELI-ME PELA VAGA DOS LIVROS DE APRESELI-ME PELA VAGA DOS LIVROS COM QUALQUER CAUSA DE UTIL E CONTINUI
 , E NEI OS LIVREINHOS, FU DINO EN NIN PESSOALMENTE, TAMPENI MO PERTEI O U NINPARAS AS CONTAS, O 13 VOCABULO 243 ENTREVISTA 243 VAGA [04]
 NOS LINDS DA VERON NA MEDIDA QUE SE ENCONTRO COM QUALQUER CAUSA DE UTIL E CONTINUI, MOI ENCLUIR NESSAS EPICODS AV ALIUPAS, NA
 ENTREVISTA E LIS EMPREGADOS, FAMILIAS E PESSOAS AMIGAS, TAMPENI TERZ FOSSE UN ESTABO QUALQUER NADO, ISTO COM ENCLUIR A SU E
 INTERAO, NES PESSOAS AMIGAS TERSE QUE VENHA ACOMPANHADO O USO [04]
 PORQUE ISTO NS PROVINCIALI AIUDA ESTI, MUITO ATRASADO, PORQUE TADO NAI AIUDA TIRADA SUFICIENTE, NAO EI, MAI MAIS FACILIDADES DE
 CONTINUI O 6 VOCABULO 144 ENTREVISTA 244 TRANSPORTES [04]
 E ACESSO A CASA F AIUDA NAO ESTAI BEM APREHECIDO PARE ISSO, EI SUO GENEIRO DE CAPEI, UM PASSATEMPO, PRIS, E QUANDO NAI QUALQUER
 ESTAI, ESTI COM UN ASPECTO MARAVILHOSO; FOI FEITO POR UN DECORADOR DE (LISBOA); PLEUS EMPREITEIROS QUE FITENNA O PRECIO; EMPREIS
 CONTINUI O 8 VOCABULO 249 ENTREVISTA 244 VAGO [04]
 S LINDAS QUE TERZ, BEI, MAS MESO SEM SER NA FEIRA NAI UM SENHOR QUE TER UMA CASA, MESO NA RUA DE (SANTIAAGO), FACEIS: MAS POR, EM
 CONTINUI O 5 VOCABULO 373 ENTREVISTA 244 TRANSPORTES [04]
 , E GIBOSE FACEISI PORQUE AS DISTANCIAS AIUDA SAO PEQUENAS PORQUE EM TRANSPORTES COLECIVOS NAO TEMOS NADA, NAO TEMOS MESO NA
 MESO NA RUA DE (SANTIAGO), FACEISI MAS POR, EM TRANSPORTES, EU DISSO FACEISI PORQUE AS DISTANCIAS AIUDA SAO PEQUENAS PORQUE EM
 CONTINUI O 6 VOCABULO 385 ENTREVISTA 244 TRANSPORTES [04]
 COLECIVOS NAO TEMOS NADA, NAO TEMOS MESO NADA, EU REFERI-ME A NIN PORQUE TEMO MEIO PROPRIO, O CARRO, NAO EI, MAS EM TRANSP
 CONTINUI O 7 VOCABULO 409 ENTREVISTA 244 TRANSPORTES [04]
 SPONTES COLECIVOS NAO TEMOS NADA, NAO TEMOS MESO NADA, EU REFERI-ME A NIN PORQUE TEMO MEIO PROPRIO, O CARRO, NAO EI, MAS EM
 COLECIVOS NAO TEMOS NADA, NAO TEMOS NADA, O MAIS USUEL EI ANDAR A PEI, E QUER TEM CARRO ANDA DE CERCO, QUER NAO TEM RESLOCIA
 O ASS DIAS HOMAS DA TANDE, DEPOIS DE NOS TEROS LAVADO, NAO TINHOS COME E LAI PARA AS CINCO HORAS «ELA GOSTAVI PULO DE ARMOS
 CONTINUI O 41 VOCABULO 124 ENTREVISTA 245 VINHA [04]
 FAZER AROZ, COMI AROZ; DEPOIS COMPREVANDI UMA SOPA SIMETICAL, CONTINUI UMA SOPA SIMETICAL LAI PARA AS OITO HORAS; PASSAVAM
 O SEQUENTE; EI QUE FOUVE UMA SERIE DE MARES; MUITO VAZ E ENTAVI UMA VEZ, ESTAVAM LAI NO CAPEI A VEM TELEVISAO, ERA POR
 CONTINUI O 37 VOCABULO 317 ENTREVISTA 245 VAGA [04]
 O MEI-MOITE E RESOLVERA IR AS PNAIA E VIRAM QUE REALMENTE AQUILO ESTAI BASTANTE VAZO; MAVA UMA EXTENSAO ENORME POR LIMA DOS
 AMIGAS QUE, DE RESIO, VAMOS LAI VEM, EI ASSIM, TER, ELE, TAO SEI, SUO NAVE, SEI, SE ELE TEM O SETIMO SNO, ELE ESTAI PARA IR PARA E
 CONTINUI O 39 VOCABULO 187 ENTREVISTA 246 VAGA [04]
 NAO EI, SE ELE NAO EI, ESTE ANO TEM QUE FAZER O SETIMO SNO, ELE ESTAI BEM ENCLUIZADO, NAO EI, NA REALIDADE ESTAI BEM ENCLUIZADO
 FAZER A REVISTA ASS MADONASSI, A ESSAS MADONASSI PORTUGAL E FALAR NINHOS BONTADE E A PARTIR NAI DA SPA SECONDA; A QUE ELAS
 CONTINUI O 39 VOCABULO 187 ENTREVISTA 246 VAGA [04]
 PAI DA CLIENTES, PORTUGAL SE QUINIA NA CLINICA E COIROS E TERO; EI TEMO COMO NIN E UN CATO PORQUE NAI, DE FERTO FOTO, EU

Annexe III

AÍSA, QUE O LANCHANTE DE DESPACHANTE, NADA É, O CLIENTE TAL PODE VER A NEGADORIA E JÁI PODE PEXER, MAS SOL* PORQUE PODE VER, POREI
 CONTAGEM 715 VOCABULO 40 ENTREVISTA 88 POR EXEMPLO
 A A HERRADIZ VIR MUITO AVARIADA, OU POUCO AVARIADA, É, PORTANTO, ASSI NÉLES O CLIENTE ATEL PODE TER INTERESSE EM DESPACHAR OUTRAS
 POIS, EU ACHO QUE SIM,
 CONTAGEM 716 VOCABULO 100 ENTREVISTA 100 POR EXEMPLO
 AQUELA NUTRIÇÃO TAL, EU ACHO QUE É UM BOM EXEMPLO DE SALVADOR NACIONAL, AQUEL INTERVENIÇÃO NELA NA TEL
 DE SE DEVEI DESTAS QUE SE DEVE CONECTAR A REPRIMIR, SEI LÁ, GREVES DE OPERARIOSI, ÉI EVIDENTE QUE EU TAMBEM ACHO QUE AS GREVES,
 CONTAGEM 717 VOCABULO 100 ENTREVISTA 89 POR EXEMPLO
 A ESSA GREVE DOS TRANSPORTES, QUE FOI UMA COISA QUE ÉI EVIDENTE QUE AFECTA MUITO A CIDADE, MAS, DE MODO MENHUR, SE OS, SE ELAS REA
 QUE ESPERAM, EN QUE SE DISSE QUE ERA NITIDAMENTE ELEMENTOS DO ANTIGO REGIME, PORTANTO, QUE INCITAVAM OS TRABALHADORES AS GREVE,
 PRAIJA DE BOM DIA, MAS SEI, NÃO SEI, MAS SEI, MAS SEI, MAS SEI, MAS SEI, MAS SEI, MAS SEI, MAS SEI, MAS SEI, MAS SEI, MAS SEI, MAS SEI, MAS SEI,
 E NADA AÍSO TANTO DE AGIM, PORQUE AGORA NADA SINAL O PROPRIO MOMENTO, O MOMENTO IDEAL PARA ELAS AGIREM, MAS, DE QUALQUER MANEIRA,
 CONTAGEM 718 VOCABULO 437 ENTREVISTA 89 POR EXEMPLO
 A PELO QUE TEI, PERO MENOS A NIVEL INTELIGENCI ACHO QUE INDA DEVE HAVER MUITAS PESSOAS, QUEI, PORTANTO, TADA PERFEITAMENTE ENQUAJ
 TACA PERFEITAMENTE ENQUAJARADA DENTRO DO PENSAMENTO FAUSTIA QUE DORINDO, PORTANTO, QUEM DIZER, NADA DIZO, NADA PENSO, PERO PEÇOS
 CONTAGEM 19 VOCABULO 174 ENTREVISTA 89 POR EXEMPLO
 NADA POUCO DIRECTAMENTE, PORTANTO, NIMA INTERVENIÇÃO, NIMA, NUN DOULQUER COISA DEI VINDA DAS DIREITAS, MAS, DE QUALQUER MODO PORN
 TEI, MAS ÉI* MOLDES PARA PLASTICOS, QUER DIZER, MOLDES DE AGORA, E DEPOIS VEMOUM-SE ASSI FABRICAS DE PLASTICOS* UM MOLDE, POIS DAI
 CONTAGEM 719 VOCABULO 175 ENTREVISTA 89 POR EXEMPLO
 PARA CIRCO-NIL, OU O QUE QUISEREM, NADA ÉI, PORTANTO OS INDUSTRIAIS, LÁI HÁI BASTANTE SIM, QUE EU SAIBA, DE BORMACHAS, QUER DIZE
 ICHREM, COMO EU EL SEMPRE AGUI, LÁI, NADA ÉI, NADA CONECTOS BEM, MAS PERO MENOS UMA SEI QUE HÁI MAS*, EN (OLIVEIRA DE AZEITEI),
 MEL PERO MENOS QUATRO, CINCO, SEIS, NADA, EU ACHO QUE CADA PESSOA TEM UM CRITÉRIO PESSOAL, DE ESCOLHER AS COISAS, NADE EIS SOL P
 O TAMPO SEMPRE A 24H, O LINDO QUE DENTRO DE MIL SAPATOS MOURMHO, HÁI MOLDS DE TAMBZ OU ATÉI OUTUMENTOS QUE NOSI GOSTAMOS, NADE ÉI, EU
 CONTAGEM 720 VOCABULO 449 ENTREVISTA 90 POR EXEMPLO
 ESCOLH, UM DESRES DOS QUE GOSTO, PARA MIM SÃO FINANÇESAS PORQUE EU GOSTO MUITO DE FRANCESZ E PORQUE ME MANDAM TAMBEM DE FRANÇA
 SIMI ISSO ÉI VENDADE, POIS MAS* ALIASI QUE SEGUIM ESTATISTICAS, NADA ÉI, MOSSAY, O NEGOCIIO DE SEGUROS NADA ÉI RE-CUSSO, ATEL
 CONTAGEM 18 VOCABULO 22 ENTREVISTA 92 POR EXEMPLO
 POR EL EN, AS COMPANHIAS DE SEGUROS AO FIM DO ANO APRESENTAM LUMOS, NADE ÉI, APRESENTAM, NADA ÉI, DUNS MILHARES DE-CENTOS, NA
 VENDADE, POIS MAS* ALIASI, QUE SEGUIM ESTATISTICAS, NADA ÉI, MOSSAY, O NEGOCIIO DE SEGUROS NADA ÉI RE-CUSSO, ATEL
 CONTAGEM 721 VOCABULO 24 ENTREVISTA 90 POR EXEMPLO
 AS COMPANHIAS DE SEGUROS AO FIM DO ANO APRESENTAM LUMOS, NADA ÉI, APRESENTAM, NADA ÉI, DUNS MILHARES DE-CENTOS, NADA ÉI, E O CE
 QUALQUER MANEIRA CONTINHO A CHER QUE D AND DE-SETES-QUATRO SÓI VAI DUM PREJUIZO, NO HANO AUTOMOVELI, TOO CONVENCIO QUE SIM,
 CONTAGEM 19 VOCABULO 347 ENTREVISTA 90 POR EXEMPLO
 O HAN, PORQUE SÓI O RACIONAL EN LUGAR DO RANOS, ÉI UM DOS BONS BANDS QUE AS COMPANHIAS TEM, NA
 A NADA HAI COM HAI IMPRESSAO DAS COMPANHIAS DE MANEIRA REHIMA, NADA HAI, SIM, EN PRINCIPIO O SEGUIM DE VIDA ÉI UMA COISA* UCH
 CONTAGEM 85 VOCABULO 484 ENTREVISTA 92 PARA JAI
 TEM UM-CE-TEZ*, TEMOS UMA CENTEZAI QUALQUER CONTATO QUE SE FAZ MOVU ÉI UM SIMISTRU QUE HAI, NADA ÉI, HÁI TARDU OU HÁI CERD, P

Annexe IV

Extrait de la liste finale du Portugais Fondamental

PAÍS	PASSADO, adj.
PAISAGEM	PASSAGEM
PAIXÃO	PASSAPORTE
PALAVRA	PASSAR
PALAVRAS-CRUZADAS	PÁSSARO
PALMEIRA	PASSEAR
PANELA	PASSEIO
PANO	PASSO
PÃO	ao passo que
PAPA, sub. m.	PASTA-DE-DENTES
PAPEL	PASTAR
PAPELARIA	PASTEL
PAR	PASTELARIA
PARA	PASTOR
para que	PATINAGEM
PARAGEM	PATO
PARAR	PATRÃO
PARDAL	PAU(S)
PARECER, v.	PAZ
parece	PÉ
PAREDE	ao pé de
PARENTE	a pé
PARQUE	PEÇA
PARTE	PEDIR
a maior parte de	PEDIR EMPRESTADO
PARTICIPAR EM	PEDRA
PARTICULAR	PEDREIRO
PARTIDA	PEGAR
PARTIDO, sub;	PEITO
PARTIR	PEIXARIA
a partir de	PEIXE
PÁSCOA	