Frequency Patterns of Phonological Objects in Portuguese - Research and Applications
PTDC/LIN/70367/2006
Duração: 36 meses

Unidade de Investigação
Principal Research Unit< /FONT< b>

Centro de Estudos de Linguagem (Onset/FL/UL)

Faculdade de Letras da Universidade de Lisboa, Alameda da Universidade
1600-214Lisboa

5. Equipa de investigação

5.1 Lista de membros (8)
5.1. Members list (8)

Nome Name	Função Role	Grau académico Academic degree	%tempo %time
Sónia Marise de Campos Frota	Inv. Responsável	DOUTORAMENTO	25
Marina Cláudia Pereira Verga e Afonso...	Investigador	DOUTORAMENTO	15(+10)
Fernando da Assunção Martins	Investigador	DOUTORAMENTO	10
Susana Mesquita de Deus Correia	Investigador	MESTRADO	10
Teresa da Costa	Investigador	MESTRADO	10
Ana Lúcia da Silva Dias Gonçalves dos...	Investigador	DOUTORAMENTO	5
Maria João dos Reis de Freitas	Investigador	DOUTORAMENTO	5
Flaviane Romani Fernandes	Investigador	DOUTORAMENTO	5

2 bolseiros a contratar

Consultants

Nome	Instituição	Fase do projecto	Custo envolvido (€)
Ana Isabel Mata da Silva	Universidade de Lisboa	Beginning, middle and final phase	

Justificação

Ana Isabel Mata has done important work on educational linguistics, including the construction of teaching materials. Her advice will be extremely helpful to attain the goals of task 4 of this project. As she is a local researcher, there is no requested funding for this consultant.

Nome	Instituição	Fase do projecto	Custo envolvido (€)
Joan Bybee	University of New Mexico	beginning, middle, final workshop	

Justificação

Joan Bybee's work is a leading reference on the studies of linguistic frequency and its role in the organization of language and the emergence of linguistic structure. She has also done specific research on phonology and language use, looking at the effects of token and type frequency. Further, she has experience in the creation of databases. Although J.Bybee is expected as a consultant, this requires confirmation closer to the project beginning.

Nome	Instituição	Fase do projecto	Custo envolvido (€)
Marina Nespor	University of Ferrara	beginning, middle, final workshop	

Justificação

Marina Nespor is a prominent researcher on Theoretical Phonology and on the Acquisition of Phonology, with strong connections with Laboratories working on Psycholinguistics and on the role of frequency and statistics in language acquisition.

Resumo (em inglês)
Abstract (in english)< /FONT< b>

Information on the frequency of phonological objects in the languages of the world, Portuguese included, is still scarce. This kind of information is relevant for fundamental research on areas such as the phonology of particular languages, linguistic universals/general trends, acquisition and development of phonology, L2 learning, and bilingual studies. Frequency is also an obvious source for establishing regular and deviating patterns in language use. Knowledge of such patterns is essential in domains like the diagnostic, evaluation and therapy of deviant/pathological speech, evaluation of proficiency of first and second language learners, or forensics. This project aims at contributing to this field: (i) by continuing the development of FreP, an electronic tool that allows the automatic extraction of frequency information of phonological objects in (European) Portuguese from written text; (ii) by building a database of frequency information of phonological objects in different types of corpora, which may then be used as reference information; (iii) by exploring the use of FreP, and the frequency data made available by the tool, for the teaching/learning of Portuguese L1 grammar; (iv) by using the knowledge of frequency patterns

in fundamental research on adult phonology and on language acquisition and development; (v) by exploring the use of FreP and the frequency data made available for forensic research; (vi) and by promoting new versions of FreP, and making the tool available to public use.
This project is a natural outcome of an informal project started in 2004, involving M.Vigário, F.Martins and S.Frota. These two years were a trial phase in which we started the development of FreP. The tool allows, at this point, the automatic extraction (identification+count) of the following units: (i) prosodic words (total number; size); (ii) clitics (total number; size); (iii) syllables (types; types by position within prosodic words; by stress condition; by position and stress condition); (iv) major class segments. Frequency information on the distribution of word stress and the direction of phonological cliticization is also obtained. To our knowledge, no electronic tool was previously available that allowed these operations. At the same time the tool was being built, the results it allowed begun to be integrated in fundamental research on adult grammar and language acquisition and development (among the topics addressed are the constraints on word size, clitic placement, the shape of early words, the emergence of syllable types). The results of this trial phase have been presented at several meetings and a web page has been built (http://www.fl.ul.pt/LaboratorioFonetica/FreP). The current (beta) version of the tool was also supplied to a number of users for research purposes.
The present project has two chief goals: the development of each of the facets of the work previously started, and the implementation of new areas of action. The following extensions of the tool are planned: (i) identification and count of all sub-classes of segments (obstruents, plosives, fricatives…), of all phonetic features (labial, coronal, dorsal, back…), considering the position within the word/syllable, and the presence/absence of word stress; (ii) frequency of phonetic segments, considering the position within the word/syllable and the status relative to word stress; (iii) frequency of word forms, by prosodic status (prosodic word/clitic); (iv) adaptation of FreP to Brazilian Portuguese. New interventions on FreP will also naturally follow from a systematic evaluation of each of its functionalities, and the identification and implementation of new procedures to avoid errors.
The information made available by FreP will lead to the establishment of a unique database of frequency of phonological objects in Portuguese. This information will be used for research in several areas: language acquisition and development, specifically of place of articulation features, of syllable constituents and of prosodic word shapes; forensic research (namely to detect speaker-particular rates of use of phonological objects, when compared to the probability rates found in the language that will be established resorting to the new database); fundamental research on adult phonology, namely in what concerns the relation between lexical phonology and actual productions, and also in comparative studies of European and Brazilian Portuguese; Portuguese L1 teaching/learning contexts, by developing materials based on the use of this new electronic tool.

Descrição dos Objectivos do Projecto
Project Objectives (description)</FONT< b>

The present project is a major contribution to the knowledge of the frequency patterns of phonological objects in the Portuguese language. Such knowledge is known to be relevant for many areas of fundamental research and its applications. The project has 3 chief goals: (1) the development of the FreP electronic tool for the automatic computation of frequency information; (2) the creation of a database of frequency patterns of phonological units in different types of corpora, to be used as reference information; (3) the application of both the FreP tool and the database to (i) research on adult phonology and the comparison of European (EP) and Brazilian Portuguese (BP), (ii) language acquisition and development, (iii) forensic research, (iv) teaching/learning of Portuguese grammar. Besides the new insights into adult and child speech, or into language variation, the project will have as outcomes new language resources for EP and BP, namely the extended FreP tool, the frequency patterns database, and teaching/learning electronic materials.

Descrição do Estado da Arte
State of the Art (description) </FONT< b>

The importance of frequency information has long been noticed in the linguistic literature. However, only recently there has been a growing interest in the frequency of grammatical objects (e.g., Beckman & Edwards, 2000; Bybee, 2001; Bybee & Hopper, 2001; Pierrehumbert, 2002; Moates, Bond & Stockmal, 2002). Frequency is argued to play a role in phonetic reduction, in the regularization of irregular paradigms, in the activation of constraints, or in the emergence and development of grammatical units such as the syllable and the word (e.g. Jurafsky, Bell & Girand, 2002; Vigário, 2003; Demuth & Johnson, 2003; Prieto, to appear). Despite the increasing interest in this domain, information on the frequency of use of phonological units in Portuguese is still scarce, and in most cases non-replicable and corpus dependent. Some frequency results have been reported for syllable structure (Andrade & Viana, 1994; Vigário & Falé, 1994, Viana et al., 1996), segments (Viana et al., 1996), word stress distribution (Viana et al., 1996), or phonetic reduction (Frota 2000, Vigário 2003). However, the data reported are either the result of manual counts or (semi)automatic procedures using tools integrated in speech synthesis systems and thus not freely available; moreover, most reports are based on the same corpus (Português Fundamental, cf. Bacelar, Marques & Segura da Cruz, 1987); and information on the frequency of several phonological objects is still absent. To fulfil this gap, an informal (not funded) project, which was the precursor of the present one, set out in 2004, involving M. Vigário, F. Martins and S. Frota.
The project consisted primarily on the development of FreP, an electronic tool for public use that now allows the automatic extraction (identification and count) of the following units, from Portuguese written texts: (i) prosodic words (total number; size); (ii) clitics (total number; size); (iii) syllables (types; by position within prosodic word; by stress condition; by position and stress condition); and (iv) major class segments. Frequency information on the distribution of word stress and direction of phonological cliticization is also available with this version of the tool. Preliminary evaluations show reliability results above 99.0%. At the same time the tool was being built, the results it allowed begun to be integrated in fundamental research on adult grammar and on language acquisition and development. Among the topics addressed are the activation of constraints on word size, clitic placement change, the distribution of syllable types, the shape of early words, and the emergence of syllable types (Vigário et al. 2005, 2006a,b; Frota et al. 2006; http://www.fl.ul.pt/LaboratorioFonetica/FreP – see the references in section 8.7 below).
To our knowledge, there is no tool similar to FreP, for Portuguese or for other languages. Additionally, the amount of knowledge that is made available with the present version of FreP and that will be attainable with the new extensions planned is also not found elsewhere: e.g., for both major varieties of Portuguese (European and Brazilian), the frequency of (i) all subclasses of segments (obstruents, sonorants, plosives, fricatives, …), considering the position within the word/syllable and

word stress, (ii) all phonetic features (labial, coronal, dorsal, back, nasal …), considering the position within the word/syllable and word stress, and (iii) all phonetic segments, also considering the position within the word/syllable and word stress. Moreover, there are no comparable data available on the frequency of phonological objects in different types of corpora (e.g. adult, child directed speech and child speech; different speech styles, etc), and reference data on the frequency of use of phonological objects in Portuguese is still missing. This bulk of information, once available, will be useful to many areas in fundamental research and its applications, such as the phonology of adult grammar in different varieties of Portuguese, language acquisition and development, forensic research, or teaching and learning of Portuguese. Its relevance to the first two domains mentioned has already started to be shown (see references above); its application to the latter two domains is a novelty, for Portuguese.

The contribution of speech science for forensics dates back to decades ago in places like the USA or UK (e.g. Nolan, 1997, www.iafpa.net, http://www.cus.cam.ac.uk/~fjn1/#forensics, www.-personal.une.edu.au/~hfraser/forensic.htm, http://web.bhm.ac.uk/forensic/). Nevertheless, even where systematic research with this goal has started years ago, it is recognized that the major limitation to the field is the lack of data on phonetic and acoustic variation in the general population. Besides that, language-specific knowledge, in addition to the identification of speaker-specific features, is essential. Research oriented towards Portuguese speakers' identification has not yet attracted the attention of the scientific community. By contrast, there has been an increasing demand by the national judiciary entities for linguistic reports aiming at speaker recognition. The gathering of the knowledge needed has thus become an urgent task, so that speech science can efficiently contribute to the Portuguese legal system. It should be noted that, to our knowledge, these requests have been addressed to the Phonetics Laboratory of the University of Lisbon, the only place in Portugal, as far as we know, where there has been a consistent response to such demand (first handled by R. Delgado Martins and then by one of our research teammembers, F. Martins).

Among the products available for Portuguese L1 teaching/learning, there are high quality materials addressing grammatical concepts, such as Duarte (2000), Costa & Costa (2001), Freitas & Santos (2001), which were, however, not conceived to take advantage of computational resources. As for electronic tools, to the best of our knowledge none of the existing products consider similar possibilities of manipulating phonological objects, and even less the computation of their frequencies (see, for example, http://www.priberan.pt/, http://www.portoeditora.pt/, http://www.escolavirtual.pt/loginPE/). It should be stressed that the use of information technology resources is very much encouraged in the official programs for Portuguese L1 teaching (e.g. Programa de Português, Ensino Secundário 2001/2002). As one of the goals of the present project is directed to meet this demand, we also hope to make a welcome contribution in the field of education.

References

Andrade, E. & Viana, M. C. (1994) Sinérese, diérese e estrutura silábica. In Actas do IX Encontro da Associação Portuguesa de Linguística (pp. 31-42). Lisboa: APL/Colibri.

Bacelar, M. F., Marques, M. L. & Segura da Cruz, M. L. (1987) Português Fundamental. Vol. II. Métodos e Documentos, Tomo 2: Inquérito de Disponibilidade. Lisboa: INIC/CLUL.

Beckman, M. & J. Edwards (2000) Lexical frequency effects on young children's imitative productions. In M. B. Broe e J. B. Pierrehumbert (Eds.) Papers in Laboratory Phonology. Acquisition and the Lexicon. Cambridge: Cambridge University Press, pp. 250-268.

Bybee, J. & Hopper, P. (2001) (Eds.) Frequency and the Emergence of Linguistic Structure. Amsterdam: John Benjamins.

Bybee, J. (2001) Phonology and Language Use. Cambridge: Cambridge University Press.

Coelho, M.C. (coordinator) (2001) Programa de Português, 10º, 11º e 12º anos. Cursos Científico-Humanísticos e Cursos Tecnológicos. Ministério da Educação, Departamento do Ensino Secundário.

Costa, J. & A. Costa (2001) O que é o Advérbio. Colecção Cadernos de Língua Portuguesa, 1. Edições Colibri / APP.

Demuth, K. & Johnson, M. (2003) Truncation to subminimal words in Early French. Canadian Journal of Linguistics, 48, 211-241.

Duarte, I. (2000) Língua Portuguesa. Instrumentos de Análise. Lisboa: Universidade Aberta.

Freitas, M.J. & A.L. Santos (2001) Contar (histórias de) sílabas. Descrição e implicações para o ensino do Português como Língua Materna. Colecção Cadernos de Língua Portuguesa, 2. Edições Colibri / APP.

Frota, Sónia (2000) Prosody and focus in European Portuguese. Phonological phrasing and intonation. New York: Garland Publishing.

Jurafsky, D. A., Bell, A. & Girand, C. (2002) The Role of Lemma in Form Variation. In C. Gussenhoven & N. Warner (Eds.) Laboratory Phonology 7 (pp. 3-34). Berlin/New York: Mouton de Gruyter.

Moates, D. R., Bond, Z.S., & Stockmal, V. (2002) Phoneme frequency in spoken word reconstruction. In C. Gussenhoven & N. Warner (Eds.), Laboratory Phonology 7 (pp. 141-169). Berlin/New York: Mouton de Gruyter.

Nolan, F. (1997) Speaker recognition and forensic phonetics. In: W. Hardcastle & J. Laver (Eds.) A Handbook of Phonetic Sciences. Oxford:: Blackwell.

Pierrehumbert, J.B. (2002) Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), Laboratory Phonology 7 (pp. 101-139). Berlin/New York: Mouton de Gruyter.

Prieto, P. (to appear) Early prosodic word acquisition in Catalan. Language and Speech (Special Issue on the Acquisition of the Prosodic Word, guest-edited by K. Demuth).

Viana, M. C., Trancoso, I. M., Silva, F. M., Marques, G., Andrade, E. & Oliveira, L. C. (1996) Sobre a pronúncia de nomes próprios, siglas e acrónimos em Português Europeu. In Actas do Congresso Internacional sobre o Português, Vol. III (pp. 481-517). Lisboa: Colibri/APL.

Vigário, M. & Falé, I. (1994) A Sílaba no Português Fundamental: uma descrição e algumas considerações de ordem teórica. In Actas do IX Encontro da Associação Portuguesa de Linguística (pp. 465-477). Lisboa: APL/Colibri.

Repercussões (descrição)

Repercussions (description)< /FONT< b>

The main outcomes of this project will be a new FreP, an electronic tool that identifies phonological objects from written text and computes their frequency (extended with additional functions, improved due to the systematic evaluation of all its outputs, and including a version for Brazilian Portuguese), and a database with the frequency patterns of phonological objects in Portuguese, including reference values for EP and BP, mean frequency values for specific groups of population (by age, sex,

profession) and different types of corpora. Both these language resources are expected to have an impact on various areas of the study of the Portuguese language, whether on fundamental research or on its applications. In the domain of the former, these tools will contribute to a better understanding of the relation between speech and grammar, and between input and output, thus providing new insights on major questions in Linguistics like how grammar evolves and how grammar is acquired. In the domain of the latter, the applications are multifarious with important impacts on (i) language teaching and learning, both of Portuguese as a first or second/foreign language, (ii) work within forensic linguistics for Portuguese, (iii) the diagnostic, evaluation and therapy of deviant and pathological speech, among many other areas.
It should be stressed that not only the tools but also fundamental research and some applications will be developed within the project, thus strengthening its impact on the national and international research communities and on language professionals. The ability to use FreP both with European and Brazilian Portuguese is a major asset, whether in terms of compared research on language variation, or in terms of potential users of the tool. The application to forensic research is a novelty with implications for national security, as to our knowledge there is no fundamental investigation, nor specific tools developed in this field for Portuguese. The new electronic materials to be developed as tools for the teaching/learning of Portuguese will contribute to the dissemination of the Portuguese language and to the assessment of proficiency in Portuguese language courses.

TAREFAS

1 - Developing FreP - adding new functions to the electronic tool: Coordinated by Marina Vigário
2 - Systematic testing, evaluation and improvement of the tool: Coordinated by Sónia Frota
3 - Creation of a database of frequency information for phonological objects in several types of corpora: Coordinated by Sónia Frota
4 - Using FreP as a teaching/learning instrument: development of new materials: Coordinated by Marina Vigário
5 - Fundamental research on the role of frequency in language acquisition: Coordinated by Sónia Frota
6 - Fundamental research on the role of frequency in adult phonology: Coordinated by Marina Vigário
7 - Application of FreP to forensic research – towards objective measures for speaker recognition: Coordinated by Marina Vigário and Fernando Martins
[Aplicações para avaliação, diagnóstico e terapia: coordenação de Sónia Frota]

TAREFA 1

Descrição da tarefa
Task decription

Designação da tarefa
Task denomination

Developing FreP - adding new functions to the electronic tool

Membros da equipa de investigação que participam nesta tarefa
Members of the research team who participate in this task

Nome Name
Sónia Marise de Campos Frota
Marina Cláudia Pereira Verga e Afonso Vigário
Fernando da Assunção Martins
Flaviane Romani Fernandes

Data de início Start date	Data de fim End date	Duração (em meses) Duration (in months)	Pessoas * mês nesta tarefa Person * month in this task
01-03-2007	28-02-2010	36	5

Resultados esperados
Expected results

Two main results are expected.
(1) The expansion of the FreP tool, integrating the ability of performing the following new operations: (i) identification and count of all sub-classes of segments (e.g. sonorants, obstruents, plosives, fricatives…), and of all phonetic features (e.g. labial, coronal, dorsal, back, nasal…), and this considering the position within the word/syllable, and the presence/absence of word stress; (ii) frequency of phonetic segments, and this also considering the position within the word/syllable and status relative to word stress; (iii) frequency of word forms, and this by prosodic status (prosodic word/clitic).
(2) A new version of the tool optimized for the Brazilian variety of Portuguese.

Descrição da tarefa
Task description

This task consists of extending FreP in order to provide the user with the ability of performing the following automatic operations: (i) identification and count of all sub-classes of segments (like obstruents, plosives, fricatives, sonorants…), and of all phonetic features (like labial, dorsal, coronal, back, nasal …), and this considering the position within the word/syllable, and the presence/absence of word stress; (ii) getting the frequency of phonetic segments, and this also considering the position within the word/syllable and the status relative to word stress; (iii) getting the frequency of word forms, and this by prosodic status (prosodic word/clitic). At the same time, a new version of the tool optimized for Brazilian Portuguese (BP) will be created. The user will be able to choose the variety of Portuguese he/she wants to work with. To adapt FreP to BP the local team will work in close connection with a Brazilian phonologist from University of Campinas.

TAREFA 2

Descrição da tarefa
Task decription

Designação da tarefa
Task denomination

Systematic testing, evaluation and improvement of the tool

Membros da equipa de investigação que participam nesta tarefa
Members of the research team who participate in this task

Nome Name
Sónia Marise de Campos Frota
Marina Cláudia Pereira Verga e Afonso Vigário
(BI) Bolseiro de Investigação (Lic. ou Bacharel) 1
Fernando da Assunção Martins

Data de início Start date	Data de fim End date	Duração (em meses) Duration (in months)	Pessoas * mês nesta tarefa Person * month in this task
01-03-2007	28-02-2010	36	38,1

Resultados esperados
Expected results

Systematic evaluation of the tool by means of the detection of the error rates for each phonological object (identification and frequency).
Classification of errors by type.
Improvement of the tool through the implementation of new procedures to avoid errors.

Descrição da tarefa
Task description

The systematic hand checking of the automatic identification of all phonological objects (both the old and the new ones) will be performed, together with the identification and implementation of new procedures to avoid errors, with possible new interventions on the tool. This evaluation will be done on the basis of the manual checking of all FreP outputs on a corpus of no less than 50,000 morpho-syntactic words.

TAREFA 3

Descrição da tarefa
Task decription

Designação da tarefa
Task denomination

Creation of a database of frequency information for phonological objects in several types of corpora

Membros da equipa de investigação que participam nesta tarefa
Members of the research team who participate in this task

Nome Name
Sónia Marise de Campos Frota
Marina Cláudia Pereira Verga e Afonso Vigário
(BI) Bolseiro de Investigação (Lic. ou Bacharel) 1
(BI) Bolseiro de Investigação (Lic. ou Bacharel) 2

Data de início Start date	Data de fim End date	Duração (em meses) Duration (in months)	Pessoas * mês nesta tarefa Person * month in this task
01-03-2007	28-02-2010	36	27,2

Resultados esperados
Expected results

Establishment of a database with the frequency patterns of phonological objects in Portuguese, including reference values for EP and BP, mean frequency values for specific groups of population (by age, sex, profession) and different types of corpora.

Descrição da tarefa
Task description

This task consists on the establishment of a so far unique database of frequency information of phonological objects in Portuguese, using the information made available by FreP. This implies computing all the frequency values provided by FreP in several types of written texts. To perform this task, a survey of available written corpora for public use is necessary. A selection of the materials found will be done on the basis of predefined criteria, taking into account the major goal of having representative data for the Portuguese language and for specific groups of the population.

TAREFA 4

Descrição da tarefa
Task decription

Designação da tarefa
Task denomination

Using FreP as a teaching/learning instrument: development of new materials

Membros da equipa de investigação que participam nesta tarefa
Members of the research team who participate in this task

Nome Name
Sónia Marise de Campos Frota
Marina Cláudia Pereira Verga e Afonso Vigário
Ana Lúcia da Silva Dias Gonçalves dos Santos
Teresa da Costa
Fernando da Assunção Martins

Data de início Start date	Data de fim End date	Duração (em meses) Duration (in months)	Pessoas * mês nesta tarefa Person * month in this task
01-03-2007	28-02-2010	36	5,8

Resultados esperados
Expected results

New materials for Portuguese L1 teaching and learning, based on a detailed description of a number of activities that teachers and learners of Portuguese can perform taking advantage of FreP.
A version of these materials adapted to the general development of grammatical and writing skills (including creative writing) as quality entertainment is also envisaged.
The resulting materials will be compiled within an electronic booklet to be incorporated in the FreP tool set. A paper publication is also planned.

Descrição da tarefa
Task description

This task consists in developing materials that explore the new electronic tool FreP for the purpose of Portuguese L1 teaching and learning. This requires the creation and subsequent detailed description of a number of activities that teachers and learners of Portuguese can perform, taking advantage of FreP's outputs. Materials will take into account the profile of users in the following ways: (i) their status as L1 teachers/learners; and (ii) the academic level/age of L1 learners. Besides Portuguese-L1 teachers and learners in the context of basic instruction, secondary studies and graduation, other target users also include teachers/students in courses of Portuguese grammar and (creative) writing, teachers/learners of Portuguese-L2 and, in general, anyone interested in developing grammatical and writing skills in Portuguese. Both the materials and the tools will also be conceived to be used for quality entertainment. Pilot experiments and demos targeting the potential users of these materials will also be conducted, in particular they will be tested by Portuguese L1 teachers and students of different academic levels.
The resulting materials will then be compiled within an electronic booklet to be incorporated in the FreP tool set. A paper publication is also envisaged.

TAREFA 5

Descrição da tarefa
Task decription

Designação da tarefa
Task denomination

Fundamental research on the role of frequency in language acquisition

Membros da equipa de investigação que participam nesta tarefa
Members of the research team who participate in this task

Nome Name
Sónia Marise de Campos Frota
Marina Cláudia Pereira Verga e Afonso Vigário

Maria João dos Reis de Freitas

Susana Mesquita de Deus Correia

Teresa da Costa

Data de início Start date	Data de fim End date	Duração (em meses) Duration (in months)	Pessoas * mês nesta tarefa Person * month in this task
01-03-2007	28-02-2010	36	8,5

Resultados esperados
Expected results

New fundamental research taking into account frequency information, in the following domains: (i) acquisition and development of place of articulation features; (ii) acquisition and development of syllable constituents; (iii) acquisition and development of prosodic word shapes. Accessing the role that frequency patterns in the input play in the acquisition and development of phonology.

Descrição da tarefa
Task description

This task consists on the investigation of several aspects of language acquisition and development, taking into account frequency information. In particular, resorting to the frequency results obtained in task 3 and those already existing for Adult Speech, and using FreP to obtain new data on Child Speech and Child Directed Speech, research is conducted in the following areas: (i) acquisition and development of place of articulation features (labial, coronal, dorsal); (ii) acquisition and development of syllable constituents (like the coda); (iii) acquisition and development of prosodic word shapes (word size, prominence).

TAREFA 6

Descrição da tarefa
Task decription

Designação da tarefa
Task denomination

Fundamental research on the role of frequency in adult phonology

Membros da equipa de investigação que participam nesta tarefa
Members of the research team who participate in this task

Nome Name
Sónia Marise de Campos Frota
Marina Cláudia Pereira Verga e Afonso Vigário
Flaviane Romani Fernandes

Data de início Start date	Data de fim End date	Duração (em meses) Duration (in months)	Pessoas * mês nesta tarefa Person * month in this task
01-03-2007	28-02-2010	36	4,2

Resultados esperados
Expected results

New fundamental research in adult phonology, taking into account frequency information, focusing on the relation between (predicted) lexical phonology and actual productions that may be affected by optional processes that are ignored by FreP computations. New insights on the similarities and differences between European and Brazilian Portuguese phonology.

Descrição da tarefa
Task description

This task consists of conducting new research on Portuguese adult phonology, taking into account frequency information. Resorting to the frequency results obtained in task 3 and using FreP to obtain new data, the investigation will include topics such as the relation between (predicted) lexical phonology and actual productions that may be affected by optional processes ignored by FreP computations. This research will provide some insight on the relation between the phonology and the phonetics of the language. Once this relation is better understood, new facilities could be integrated in a tool like FreP, in particular the activation of phonetic implementation rules that would deal with some of the optional processes (e.g. unstressed vowel deletion).
Taking advantage of the new version of FreP adapted to Brazilian Portuguese, comparative work on the phonology of both major varieties of Portuguese will be conducted.

TAREFA 7

Descrição da tarefa
Task decription

Designação da tarefa
Task denomination

Application of FreP to forensic research – towards objective measures for speaker recognition

Membros da equipa de investigação que participam nesta tarefa
Members of the research team who participate in this task

Nome Name
Sónia Marise de Campos Frota
Fernando da Assunção Martins
Marina Cláudia Pereira Verga e Afonso Vigário
(BI) Bolseiro de Investigação (Lic. ou Bacharel) 2

Data de início Start date	Data de fim End date	Duração (em meses) Duration (in months)	Pessoas * mês nesta tarefa Person * month in this task
01-09-2007	28-02-2010	30	12,9

Resultados esperados
Expected results

Useful data for speaker recognition will be provided, in particular frequency values of phonological objects that can concur to identify particular individuals. A table of allowed frequency ranges for European Portuguese and the identification of outliers will be attempted.

Descrição da tarefa
Task description

The goal in this task is to compare the frequency of use of phonological objects of specific target individuals with the frequency of use of the same objects in a larger set of the population. The ultimate objective would be to find frequency values that could concur to identify particular individuals. This work requires the outputs expected in task 3, which involve the computational of mean frequency values for different sets of the population, together with the gathering of individual corpora in order to run FreP and obtain relevant data for specific individuals.

Recommendation of Funding:

75.000,00 €

Due to budgetary limitations, the funding for books, travel, and other miscellaneous expenses has been reduced, and the funding for consultants has been eliminated entirely.

Recommendation of Funding for Human Resources:

0,00 €

As pleased.

Comments to be trasmitted to the proponent:

The panel found this proposal to be one of the best applications. The project is original in its aims and has extremely high potential. It emphasizes the relevance of the knowledge of the frequency patterns of phonological objects for fundamental research in many linguistic areas (phonology, typology, language acquisition, language use, forensics, etc.). Frequency information is vital to any current learning algorithms attempting to model human cognition, and having a resource such as this on hand for future research would prove extremely valuable. Given that frequency studies are still scarce, this project, which is the continuation of an earlier (informal) project, tries to fill this gap. The main goal of the project is the development of an electronic tool for the automatic computation of frequency information about phonological objects in (European and Brazilian) Portuguese from written texts and the creation of a database of frequency patterns of phonological units in different types of corpora. It is expected that both the electronic tool and the database will "...have an impact on various areas of the study of the Portuguese language, whether on fundamental research or on its application“. The panel hopes that in communicating the results of this project both in the research community and to funding agencies, the investigators will take care to present details about how FreP is structured.