

1.	Bernhardt, Barbara.....	2
2.	Davis, Barbara .....	3
3.	Demuth, Katherine.....	4
4.	Freitas, Maria João.....	6
5.	Goad, Heather.....	7
6.	Goldstein, Brian.....	10
7.	Hamody, Anmar.....	11
8.	Kern, Sophie.....	14
9.	LeNormand, Marie-Thèrese .....	15
10.	Levelt, Clara .....	17
11.	Lleo, Conxita .....	20
12.	Cruz-Ferreira, Madalena .....	22
13.	Nunes, Aubrey.....	23
14.	Ota, Mits.....	26
15.	Pater, Joe.....	28
16.	Prieto, Pilar .....	30
17.	Ratner, Nan.....	32
18.	Schneider, Katrin .....	33
19.	Smit, Ann.....	35
20.	Stoel-Gammon, Carol .....	37
21.	Van der Pas, Brigit.....	38
22.	Velleman, Shelley .....	40
23.	Vihman, Marilyn.....	41
24.	Wauquier-Gravelines, Sophie.....	42
25.	Young-Scholten, Martha .....	45
26.	Zamuner, Tania.....	46
27.	Zmarich, Claudio .....	48

## 1. Bernhardt, Barbara

**To: Brian MacWhinney and Yvan Rose**

**RE: The Phon Project for CHILDES**

The Phon project is a long-anticipated initiative in the field of phonological acquisition. I am very much interested in participating in the consortium organizing this new database.

### 1. **Broad research questions:**

- a. What are the statistically frequent and infrequent patterns in monolingual and bilingual phonological development cross-linguistically at various cross-sectional points and longitudinally for
    - i. Word structure and stress patterns
    - ii. Onset, rime, and intervocalic consonant development
    - iii. Segmental and feature development
    - iv. Sequences of elements, both contiguous and noncontiguous
  - b. What are the statistically frequent and infrequent patterns of interaction between phonological and other aspects of language development?
  - c. How do children with and without phonological impairment compare?
2. **Importance of the questions:** Large-scale studies in phonological development are beyond the scope of individual researchers, and thus, there is still a paucity of information about all aspects of phonological development from which to develop and test theories and advance clinical phonology (identification of impairment and delineation of paths for intervention). Theoretical perspectives of interest include at a minimum: interactivity versus modularity in monolingual and bilingual linguistic systems (the notion of ‘template’ in early acquisition, the representation of consonants vs vowels, and the interaction of different phonological and morphological levels), OT perspectives on acquisition (e.g., constraint demotion vs promotion), and markedness and defaults in phonological systems.
3. **Contribution:** I can contribute transcribed pre- and post-intervention samples from 60-70 English-speaking children with phonological impairment (audiotapes collected in field conditions) and 10 speakers with hearing impairment (audiotapes and ultrasound movies collected in laboratory conditions). The audiotapes and ultrasound movies may or may not be available due to research ethics restrictions.
4. **Analytic features required:** A clear process for data entry and analysis that allows:
  - a. Nonlinear phonological analyses for word, syllable, rime, onset, mora, nucleus, coda, segment, feature (single, and multiple features), sequences: inventory analyses, and comparative analyses, including within-sample comparisons between child and adult targets (for all constituents), and between-sample comparisons, including analyses between phonological and other language samples with and across children and languages.
  - b. Researcher-led syllabification and analysis selections
  - c. IPA format for data entry and presentation

Barbara Bernhardt, Ph.D. Associate Professor, School of Audiology and Speech Sciences, University of British Columbia, Vancouver, BC, CANADA, V6T 1Z3

## **2. Davis, Barbara**

December 17, 2004

I am committed to participation in the consortium organized by Brian MacWhinney and Yvan Rose. I understand that the goal of this consortium is to make existing phonological databases available in a consistent format for research purposes. I would like to emphasize my support for the proposed project as a means to make high quality acquisition data available to interested researchers. As a consultant on the proposed project I will work each year for three months over the summer preparing the Texas and the Stanford corpora, testing the software for running analyses, and developing new analyses for research questions pertinent to the proposed databases.

My general research goals which are relevant to the proposed database center on understanding the emergence of phonological knowledge from phonetic abilities available to young children. I study typical speech acquisition with an emphasis on cross-linguistic comparisons to understand common patterns and patterns unique to either individual children or to learning from the ambient language. I also am interested in translation of perspectives from study of typical acquisition to understanding children with developmental speech delay or disorder. Particular clinical populations of interest include hearing impairment (with emphasis on children identified at birth with varying degrees of hearing loss and children who receive early cochlear implantation), "late talking toddlers", and children with severe levels of speech delay or disorder. In this regard, I would like to have access to acoustic analysis as well as perceptual transcription based data beginning at the onset of canonical babbling and extending until the acquisition of ambient phonology (approximately 7 months until 6 years of age). I am also interested in analysis of Infant-Directed and Adult-Directed speech styles employed in language directed to children who are 7 month-24 months of age.

I have included an appendix which details some of the precise analyses which I might be interested in pursuing with the proposed database. I have detailed general characteristics, populations of interest, and categories of research questions.

Data that I could contribute to this project includes audio data on 15 children during canonical babbling, 18 children during the single word period, and 12 children during the period between 24 and 36 months of age. All are from monolingual English environments. The children were audio recorded weekly during babbling and monthly during the single word period and beyond. The data are presently available in LIPP format.

Sincerely,

Barbara L. Davis  
Professor

### 3. Demuth, Katherine

October 14, 2005

Brian MacWhinney  
CMU - Psychology  
5000 Forbes Ave.  
Pittsburgh, PA 15213

Dear Brian MacWhinney,

As you may know, we have been involved over the past 4 years in a large crosslinguistic project to examine children's acquisition of phonology and morpho-phonology between the ages of 1-3. We are particularly interesting in understanding how children's phonological grammars develop over time, if they show sensitivity to frequency effects in the ambient language, and if interactions at the phonology/morphology interface can shed some light on the variable production of grammatical morphemes. All of these issues are relevant for a greater understanding of the mechanisms of language learning and the course that this takes in children with normal language development. It will also lay the normative baseline for investigating similar issues in bilingual and language impaired populations.

These questions need to be addressed in crosslinguistic perspective, but this has been difficult to date due to the lack of phonetically transcribed data on any language available in the public domain. We are therefore very excited to know that, as part of this project, you are identifying other researchers who will also be donating their phonologically transcribed materials. Only in this way can the field truly move forward, enabling us all not only to ask, but begin to answer many of the research questions that to date have been impossible to address.

As part of our current research we have therefore been collecting longitudinal data from 6 English-speaking and 6 French-speaking children between the ages of 1-3, audio- and video-tapping them every two weeks. We are transcribing parent and child utterances according to CHILDES format, including broad phonetic transcription of the children's speech. Once this is completed, we plan to donate the *Providence Corpus* and the *Lyon Corpus* to the CHILDES database, providing a valuable set of transcripts for other researchers. We have already sent you 2 of the files, and the other 10 will be sent by June, 2006.

Of course, having the data is one thing. Having the tools to help analyze them is another. Tools for carrying out phonological analysis have been seriously lacking, forcing

researchers to resort to extremely time consuming, labor intensive searches by hand. In conjunction with colleague Mark Johnson we have therefore been trying to develop some research tools of our own. However, rather than having everyone reinvent the wheel, it would be ideal to develop a set of standard tools that could meet most researchers' needs. This toolbox would ideally include the ability to:

1) Match children's (and adults') target words to and IPA rendition of the target, using an on-line dictionary. This would provide the IPA for the target word.

2) Provide a probabilistic alignment of the target word IPA with IPA child word production on the %pho tier, segmenting it from running unsegmented production forms (the form in which our data, and many others (e.g. the Deuchar corpus) is transcribed)

Items 1) and 2) are critically important, and probably the most challenging to program.

3) Pull out all target words of a certain shape, based on number of syllables, position of stress, etc.

4) Pull out all target syllables of certain shapes, including complex onsets, complex nuclei, complex, complex and simple codas/word-final elements, etc.

5) Provide syllabification options for both target words and children's IPA productions (stress attracts vs. no-stress attracts (e.g. CVC.V vs. CV.CV).

6) Provide segmental inventory, as well as contexts (syllable/word-initial, syllable/word final, syllable-final word internally) in which segments occur.

It would be helpful to have items 3)-6) extracted, but also displayed on the same spreadsheet with their larger sentential context, so that contextual effects can also be adequately assessed and coded for without having to cut and

Finally, having all extracted/paired forms displayed in spreadsheet format, or easily portable into one, would greatly facilitate coding of the data.

I am therefore extremely interested in participating in the consortium that will be involved in organizing this new phonological database.

Sincerely,

Katherine Demuth

#### 4. Freitas, Maria João

Dear Prof.s Brian MacWhinney and Yvan Rose,

This is a formal letter stating that the following Portuguese researchers

Maria João Freitas  
Teresa Costa  
Susana Correia  
Letícia Almeida

working on phonological development at the Departamento de Linguística Geral e Românica, Faculdade de Letras, Universidade de Lisboa, Portugal, accept your proposal of participating in the consortium organizing the phonological database *Phon Project*. The researcher Teresa Costa is already using a previous version of the database; Susana Correia and Leticia Almeida are now starting to use the available format of the database to proceed with their research.

The research topics we are now addressing are (i) the acquisition of syllable structure, (ii) the acquisition of place of articulation, (iii) the acquisition of word stress, (iv) the acquisition of the vowel system in Portuguese/French bilingual children and (v) the effects of phonological processes in the structure of early lexical representations.

So far, we have collected longitudinal data from 8 monolingual Portuguese children (2 children have been videotaped for 3 years; 1 child has been videotaped for 2 years; 5 children have been videotaped for 1 year). We are now in the process of data collection involving 3 bilingual children (2 Portuguese/French children and 1 Portuguese/Cape Verdian Creole child) and 2 monolingual Portuguese children.

Considering the research topics we mentioned above, we do not need very sophisticated analytic tools. As far as we know, the current version of the database accurately accounts for the analysis of syllable structure and segmental properties. We also need efficient tools to evaluate stress and to easily codify the data in order to describe the acquisition of phonological processes.

The above mentioned Portuguese researchers deeply thank Prof.s Brian MacWhinney and Yvan Rose for organizing this database for the researchers working in the field of phonological acquisition.

Lisboa, October, 12, 2005

(Maria João Freitas)

## 5. Goad, Heather

Brian MacWhinney and Yvan Rose  
Department of Psychology  
Carnegie Mellon University  
5000 Forbes Ave.  
Pittsburgh PA 15213

19 October 2005

Dear Brian and Yvan,

I am writing to express my unconditional support for and interest in the *Phon* project, including the creation of a phonological database to be shared across researchers in the field. A research tool such as this has long been necessary in phonological acquisition and will surely advance scholarship in the field due to increased visibility, productivity, accountability, and collaboration.

Visibility: The field of phonological acquisition is relatively small. This is principally due to the unavailability of phonetically-transcribed data in the public domain. The fact that one cannot simply go to the library and flip through the grammar of the phonology of child X in the same way that one can flip through a grammar of Tagalog means that the opportunity for a set of data to attract someone's attention in passing is simply not available. (For example, researchers working on Reduplication in Tagalog may find the parallels that exist between this process and Consonant Harmony in early grammars striking if they happened to come across some data on the latter topic; an examination of the similarities and differences between these two processes could serve to illuminate the analysis of both phenomena.) Similarly, while researchers who work on end-state grammars often recognize that having access to data on the acquisition of some phenomenon may serve to discriminate between alternative accounts of the phenomenon, the lack of accessible data leaves this important avenue of investigation largely unexplored. (For example, access to information on the German child's treatment of target [kn] could shed light on whether this cluster is best analysed as a branching onset, akin to [kl], or as an appendix-initial cluster, akin to [Sn].) The establishment of a shared phonological database will go a long way toward solving these problems. It will also bring new people and new ideas into the field and will increase the visibility of the work currently being done; this, in turn, will enhance scholarship in the field.

Productivity: At present, a significant portion of the research time of scholars in the field of phonological acquisition is devoted to the collection, transcription and coding of new data as well as to entering the few comprehensive sources of data available in the published literature into personal databases (e.g. the Amahl corpus (Smith 1973)). While the value of collecting new data should not be underestimated, the collection of new data for each new research topic is often necessary because of the lack of data in the public domain. The creation of a publicly-available database with the accompanying segmentation software package will naturally make this burdensome task unnecessary, freeing up significant resources to devote to analysis and interpretation. This can only push the field forward in its level of sophistication.

Accountability: At present, generalizations and analyses are published without any opportunity for other scholars to check the reliability of the data. Important generalizations can inadvertently be missed when only one researcher has access to the data from a particular child. Following from this, alternative competing analyses of the same phenomenon are often difficult, as data required for the reanalysis may simply not have been reported in the original source. This situation does not typically characterize other fields of linguistics and it severely limits the opportunities for intellectual debate and discussion. Sharing within a forum such as *Phon* will surely enhance the level of scholarship.

Collaboration: The availability of data through a mechanism like *Phon* will no doubt lead to greater collaboration among scholars, as researchers working on similar questions will have access to the same body of primary data. Since collaboration brings people with competing ideas and backgrounds together, it raises the level of debate, thereby having the potential to advance scholarship in the field. I will now turn to elaborate on my personal interests in *Phon*. I belong to a group of researchers who define themselves as theoretical phonologists who work principally with acquisition data. One of the overarching questions that guides the kind of research undertaken by this group is: how does linguistic theory constrain what a possible developing grammar is? As I will briefly illustrate below, the development of *Phon* will have a positive impact on work which falls under the scope of this question.

Since the advent of Optimality Theory, there has been significant interest in the field of phonological acquisition in drawing comparisons between the typological variation exhibited across adult grammars and the stages in development through which learners pass on their way to arriving at the target grammar. This kind of investigation will be greatly facilitated by the development of *Phon* because a comprehensive investigation of a given empirical issue within this approach requires access to longitudinal data and to cross-linguistic data. Given past data collection trends in the field, much of the data that will be contributed to *Phon* will likely be longitudinal. As well, in recent years, the field of phonological acquisition has seen an explosion of work on languages other than the most commonly studied English and Dutch. Regarding the latter, the importance of *Phon* cannot be underestimated. For example, while a researcher based in the mid-western United States may recognize that his/her analysis of some phenomenon would be greatly enhanced with data from the acquisition of Japanese, the difficulty in finding children who are learners of this language in his/her region make this avenue of investigation nearly impossible to pursue. This, naturally, could be resolved through a shared phonological database.

Much of my recent work has been on syllabification. One area of typological variation that I am particularly interested in exploring concerns the syllabification of sibilant-initial clusters. Some languages do not permit such clusters at all (Spanish) while others only permit obstruents to follow the sibilant (French, Greek). Some only permit [s] in cluster-initial position (English, Dutch, French), others only permit [S] (German), and others are more permissive ([s] and [z] in Greek; [s], [z], [S], and [Z] in Russian). Finally, languages differ in what prosodic constituent licenses the sibilant (the syllable in English and Dutch; the prosodic word in German). While there is quite a large literature on sibilant-initial clusters in acquisition, most of it focusses on English and all of it, to my knowledge, focusses on the similarities and differences that hold in the development of these clusters as compared to regular branching onsets. None of the typological differences mentioned above have been investigated in development, a topic which again requires access to longitudinal data and to cross-linguistic data, precisely the kind of data that will be available through *Phon*.

A large component of my research involves arguing that representations in

syllabification are highly-articulated. This view, however, is not shared by everyone; in fact, there is a commonly observed trend in the field against this position. In light of this difference, I would like to comment on the analysis capabilities of *Phon*. I am particularly interested in the goal to make the segmentation software as theory-neutral as possible and, at the same time, to allow some flexibility on the part of the researcher. To my mind, this is important for two reasons. One, it allows individual researchers who do not share standard assumptions about syllabification to assign structure to segments in particular positions as he/she sees appropriate; two, it enables the researcher to test alternative analyses through various parses of a string of segments. Of utmost importance, in the preliminary versions of *Phon* that I have seen, this type of flexibility does not come at the expense of user-friendliness!

Naturally, the success of *Phon* depends on the willingness of researchers in the field to contribute corpora to the shared database. I would like to contribute the following corpora:

1. Clara: L1 Québec French; 1;00,28-2;07,20; spontaneously-elicited (Goad & Rose shared data);
2. Théo: L1 Québec French; 1;10,26-4;00,00; spontaneously-elicited (Goad & Rose shared data);
3. Hannah; L1 Canadian English; 2;00,01-2;05,18; spontaneously-elicited; 4. David and Mark (twins); L1 Canadian English; 3;3-3;7, spontaneously-elicited.

In closing, I am very interested in participating in the consortium organizing this new phonological database. I think the work is immensely important and I wholeheartedly support it.

Sincerely,

Heather Goad  
Associate Professor

## 6. Goldstein, Brian



Brian MacWhinney  
CMU - Psychology  
5000 Forbes Ave.  
Pittsburgh, PA 15213  
Dear Dr. MacWhinney,

It is my pleasure to provide a letter of support for the PhonBank project. This project will be logistically, theoretically, and clinically important. Existing databases are generally much more facile for morphosyntactic analyses than for phonological analyses. Although the CHILDES database has been enormously successful for morphological and syntactic analyses, completing phonological analyses has been significantly more cumbersome. This has made it difficult for researchers to share data and collaborate on projects, tasks that are much more facile for other areas of language. PhonBank has the potential to obviate the limitations of phonological analyses that now exist through other databases (e.g., means of data sharing, customizing data fields, merging transcriptions, and segmenting phrases). In addition, including lines for IPA targets and the child's production are further assets of the program. It is my hope that the full set of IPA symbols are included along with the IPA extension symbols that will allow accurate transcription of the speech of typically developing children and children with phonological disorders. Another significant improvement of PhonBank over other systems is the ability to syllabify. On other platforms, the ability to syllabify is non-existent, limited, and/or awkward. PhonBank attempts to solve these issues through parameters that are user-definable. These improvements over other existing systems will allow researchers to ask questions and mine transcripts in ways that were not previously possible. Added to these functions are ones that incorporate statistics and charts. Thus, PhonBank has the potential to be the most powerful tool phonologists have to store, code, and analyze data. I hope to contribute samples of monolingual Spanish-speaking 3- and 4-year-olds and bilingual 4- to 6-year-olds. The majority of these children are typically developing although many exhibit phonological disorders. I hope to use the database to examine the issue of phonological representation in bilingual children. To do this, it is imperative to examine cross-linguistic aspects of phonological acquisition. This type of research is only practical with the type of database that is being proposed in this project.

In summary, I am quite enthusiastic about this project and am willing to participate in the consortium building this new database. I also look forward to using it in my own work. Please feel free to contact me at 215-204-7593 or at [briang@temple.edu](mailto:briang@temple.edu) if you have any questions.

Sincerely,  
Brian Goldstein, Ph.D.  
Associate Professor

## 7. Hamody, Anmar

Dear Professor Rose,  
Below I reiterate my support to the Phon system.

Assuming that consonant clusters constitute a source of difficulty to young children, we have chosen to address the following questions: First, how do early productions of Arabic children deviate from those of the adults' with respect to words having onset clusters ? Second, what are the simplification strategies implemented by Arabic children when using words with clusters ? Third, if cluster reduction is a well-attested strategy in cluster acquisition, then which consonant(s) of the cluster is deleted ? Fourth, what are the stages of development that demarcate the child's progression in cluster acquisition ? Providing answers to these questions helps give insight into one aspect of phonological acquisition, viz.the acquisition of consonant clusters by Arabic children.This study is mainly theoretical, yet it may provide useful background to subsequent research that is planned to have clinical or educational applications.To carry out this study, 13 Arabic-speaking children (aged 2-3 years ) have been taken as subjects.The data analysed in the study represent the spontaneous utterances of these children, accumulated longitudinally along a period of 9 months.Thus, the corpus of our study constitutes about 2200 tokens of words containing onset clusters.

From the introduction to the Phon system offered on the website, I noticed that this system provides a means for analysing child speech.Thus, among the facilities it provides is one that relates to analysing the syllable structure of the words produced by children.This in itself helps me, as a researcher, to analyse my data easily and accurately.Moreover, and more importantly, my study,as far as I know, is the first one dealing with the acquisition of Arabic consonant clusters and to conduct such a study supported by computational analysis will be a contribution to the literature of child phonology.

I am interested in participating in the consortium organizing the Phon system.

Miss Anmar H. Sa'eed

Howell, Peter



Department of Psychology  
**UNIVERSITY COLLEGE LONDON**  
GOWER STREET LONDON WC1E 6BT

Tel: 0207 679 7566  
Fax: 0207 436 4276  
E-mail: p.howell@ucl.ac.uk

January, 2005

Dear Brian and Yvan,

The Speech group in the Department of Psychology, University College London is interested in how stuttering (a developmental speech disorder) evolves out of fluent speech. Over the last 20 years we have collected data bases on fluent children and children who stutter and many of the participants have provided multiple speech recordings at different ages (spontaneous monologue and in dialogue with a fluent interviewer) as well as a range of demographic, language-performance measures and so on. The children who stutter are differentiated into those who persist and those who recover at teenage (at this age, any child still stuttering is unlikely to recover).

The main research questions which we wish to address are:

1. To determine whether the language of children who stutter differs from their peers around the time of onset of the disorder.
2. To determine whether the language of children who are destined to persist in their stutter (established at teenage) differ from children who are destined to recover around the time of onset of the disorder.

We believe these data are important for theoretical questions concerning our understanding of normal fluent speech development and for establishing what can go wrong and lead to stuttering. At a practical level, differences between speakers who persist and recover from their stutter can potentially be used to identify how language might be changed in the speakers who persist (i.e. generate potential hypotheses about how they should be treated).

The research data we would like to contribute to the PhonBank project are speech recordings (spontaneous monolog and dialog) that are processed using CHAT, speech filing system (SFS) and PRAAT formats. A data release has been made of some of the monolog material from children who stutter (the UCLASS database which can be accessed by visiting <http://speech1.psychol.ucl.ac.uk/index.htm>). We plan to make dialog data from these speakers and fluent speakers available shortly (subject to ethics approval committee).

A number of different analyses of these data are possible given the flexibility in the Phon and other software packages. It is hoped that by contributing these data, other groups will be able to examine their own hypotheses about the questions we have raised (and other) on these data.

In my opinion, there is a clear need for PhonBank as no database currently exists such as the one envisaged to investigate the issues I have mentioned. In particular, it opens up the possibility of addressing the questions I raised from a cross-linguistic perspective (we ourselves do small-scale work on Spanish as well as on English).

As head of the Speech group in the Department of Psychology, University College, I want to state formally that we are interested in participating in the consortium organizing this new phonological database.

Yours sincerely,

Professor Peter Howell

## 8. Kern, Sophie

Sophie Kern  
Laboratory Dynamique du Langage  
Institut des Sciences de l'Homme  
14, avenue Berthelot  
69367 Lyon cedex 07  
France  
e-mail : [sophie.Kern@univ-lyon2.fr](mailto:sophie.Kern@univ-lyon2.fr)

Lyon, 29<sup>th</sup> of September 2005

My research focuses on very early language acquisition from a cross-linguistic perspective. The main goals are to describe and to explain common trends observed in babbling and during the first words period. A cross-linguistic perspective is adopted for comparing systematically babbling, first words and lexical spurt in infants acquiring typologically different languages: Dutch, Rumanian, Turkish, French, Berber and Tunisian. At least 4 children per language have been recorded in their homes every two weeks from 8 months of age through 25 months of age. No structure is imposed on the normal household routine.

This methodology has led us to a series of interesting results concerning the universals and specificities of early phonetic and phonological development. Data provide further evidence for a universal basis for vocal patterns in babbling. For example children produced more stops, nasals and glides than other types of sounds; they tended to use more coronals and labials than dorsals and gutturals. Vowel production data showed a strong preference for vowels belonging in the lower left part of the vowel space, a trend reported across many studies of this period. However, these data also seem to indicate some emergent influence from the ambient language as early as the babbling period. Other analyses on first words are under progress.

2/3 of entire corpus (canonical babbling and words) has been transcribed using broad phonetic transcription conventions and entered into a computer database using Logical International Phonetic Programs software designed for describing phonetic patterns. The LIPP software is interesting software in so far as up to now it was the unique one that allowed phonetic and phonological analyses. However, its use is time consuming, not really user-friendly and it doesn't allow for a multimedia treatment.

The PHON database program seems to be a very promising tool. In the future, we intend to look closer at the phonological but also at the semantic and grammatical nature of first words produced by children. By establishing a link between pictures and sounds, PHON will allow us to integrate contextual elements into our analysis. It will also give us the chance to run queries on segments type, on segments co-occurrences as well as on syllabic structure of first words. Moreover, as one of our aims is to identify phonological universals, we could use PHON data base in order to multiply the number of studied languages or on the contrary share our corpora with colleagues who address different kind of questions.

Two expansions of the project described above would take also advantage of the functions offered by PHON. One of the expansions concerns motherese, which will be described in phonological but also in semantic and morpho-syntactic terms. We intend to enter all data into the PHON database in order to study the role of maternal input on very early language development. The second expansion is the study of gestures in children and in adults. Here again PHON is a very useful tool which gives us a lot of contextual elements without which it is very difficult to attribute a function to a specific gestures. Moreover, gestures will be considered in relation to words and morpho-syntactic emergence.

PHON could on the one hand be a very powerful tool for data sharing and is on the other hand an effective tool for data transcription and analysis. Therefore, I would be really interested in participating to the consortium which will organize this new phonological database.

Sophie Kern

## 9. LeNormand, Marie-Thèrese



**Marie-Thérèse Le Normand**  
Directeur de recherche  
INSERM

**Hôpital Robert Debré**  
48, Bd Sérurier - Paris, 19  
**Service de Neurologie Pédiatrique**  
**et des Maladies Métaboliques**

To Brian MacWhinney and Yvan Rose,

I would like to support the PhonBank project which will greatly contribute to central issues in language development and to research programs currently conducted internationally, including at the INSERM Research Center. Of particular interest is our research program in developmental cognitive neurosciences entitled “The Neural Bases of Speech and Language Development of Children with Neurogenic Disorders”. I am actively involved in this research at the hospital for children Robert Debré, in Paris, France.

For many years already, I have used the Child Language Data Exchange system (CHILDES; MacWhinney & Snow, 2000) tools to collect free speech data from French infants and preschoolers. Many of these children had suffered from early brain damage (usually in the pre- or peri-natal periods). The learning process changes markedly when these children have to convert the same speech sound patterns into motor output. At this point, perceptual detail may be of critical importance but also early emergence of productive control over different components of language (phonology, lexicon and grammar).

Research in areas such as these requires access to large corpora of phonological data. However, phonological investigations of child language based on large corpora of data remain a challenge still currently unresolved. If it is the case that left temporal cortex plays a critical role in the extraction, storage and reproduction of perceptual detail (visual and/or acoustic), then children with early brain damage will be at a greater disadvantage in this phase of learning. However, once the required patterns are constructed and set into well-learned routines, the left temporal disadvantage may be much less evident.

In order to understand the cerebral involvement and functional plasticity in speech acquisition, the research questions I would like to address, based on the PhonBank database, are the following:

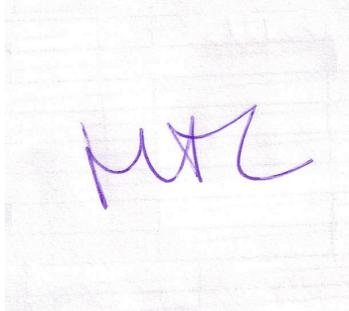
- 1) To what extent individual differences in auditory memory may be responsible for phonological variation?
- 2) What are the effects of early brain damage investigated in the first stages of phonological development? The period from first word to grammar is, for example,

the period in which we should expect to see maximal plasticity, but it is also the period in which the initial specializations of particular cortical regions ought to be most evident.

- 3) To what extent recovery of phonology could be assessed in children with focal brain injury?
- 4) How is it possible to examine left-hemisphere specialization for language in all components of language?
- 5) How to resolve apparent contradictions between plasticity and early specialization for language?
- 6) How does phonological development differ through cross-language differences?

The importance of these issues in developmental cognitive neuroscience led me to conduct follow-up studies with children with atypical speech and language development: (deaf children with cochlear implants, “consonantless” children, a form of dyspraxia, premature children, epileptic children with congenital left temporal lesion, SLI children). So far, repeated and detailed linguistic analysis in language production were undertaken overtime using the tools available on CHILDES at different developmental language stages in order to examine whether or not all components of grammar (phonology, lexicon and syntax) develop in a way strikingly similar to the patterns documented in normally developing children. Video and audio recordings were used. All sessions have been transcribed in accordance with the guidelines produced by the Codes for Human Analysis of Transcripts (CHAT), which is part of CHILDES. All these French databases are now available for phonological analysis. The particular analytic features in Phon I that I am most interested in are running media segmentation to analyze consonant productions in segmental sequences (irrespective of syllable structure) such as CC sequences (from the perspective of place and manner of articulation) and CVC sequences (focusing mostly on the place of articulation of the segments involved in the sequences).

Such a multimedia database for child phonological data storage and analysis, which automatically performs parses and calculations, based on the phonetic transcriptions of the target words and the child’s renditions of these words is very promising for research and remediation. The important feature of the system proposed as part of the PhonBank project is that the results of the data compilations are instantly generated by the computer in a reliable fashion and can be compared using large language-specific or cross-linguistic corpora. All French children involved in the Phon database will be part of a larger project supported by L’Institut National de la Santé et de la Recherche Médicale (INSERM), Paris. I am of course really interested in participating in the consortium organizing this new phonological database.



## 10. Levelt, Clara

The *Phon* project provides the long-awaited source for research on phonological acquisition. The field has always suffered from inaccessible, hard-to-find, and hard-to-gather data. With *Phon* the quality of research on phonological acquisition will be raised enormously: theories will be based on large amounts of data and cross-linguistic data, and it will become not-done to refuse to make data available for replication or reanalysis.

My research, and co-authored research with, among others, Paula Fikkert (Nijmegen University), Niels Schiller (Max Planck Institute), Willem Levelt (Max Planck Institute) Ruben van de Vijver (Potsdam University) Paul Boersma (University of Amsterdam) and Joost van de Weijer (Lund University) has, up until now, been centered around data from children acquiring Dutch. This has provided insight into, on the one hand, the development of syllable structure, and, on the other hand, Place of Articulation patterns in the development of words (see references below). Based on this research, clear hypotheses concerning development in languages other than Dutch have been formulated. These hypotheses await to be tested on cross-linguistic developmental data. Initial attempts to do this have been made (Fikkert, Levelt & van de Weijer, 2003; Fikkert, Levelt, Wauquier, Freitas & Grijzenhout 2004), but were severely hampered by either the unavailability of data, or the often unprofessional way in which the data from the different sources were stored, which made searching for particular characteristics extremely time-consuming.

The syllabic development of Dutch children - a database of 12 children was used, see below for more details - shows a very systematic pattern. Phonological theory predicts twelve possible learning paths that learners could follow from the initial state of the grammar to the final, adult, state. Dutch children turn out to follow only two of those. It appears that the input-frequency of syllable types in Dutch acts as a sign-post whenever the learner faces a choice between two possible intermediate grammars. It is obvious that we need to carefully study syllabic development in other languages than Dutch to see whether (1) developmental paths are equally mono-directional in other languages, and (2) whether input-frequency correlates with a particular developmental route in other languages as well, and/or (3) whether other sign-posts are available to the learner to guide them through the developmental possibilities.

Similar questions can be asked on the Place of Articulation side. Again, a fairly mono-directional developmental pattern was found in the data of Dutch children. This was analyzed as being related to the particular form of the earliest lexical representations, and, for later developments, to input-frequency of particular patterns, and emerging constraints relating to input frequency in the developing grammar of the learner. Again we need to know what developmental patterns are found in data from children acquiring other languages than Dutch, and whether the developmental representations hypothesized for Dutch, and frequency-information from the language-specific input can account for the cross-linguistic data as well.

*Phon* will be tremendously helpful in carrying out these research projects. First of all, several people have promised to contribute their data to *Phon*, and these data are from a

wide variety of languages and language families. Second, *Phon* has a host of analytical tools that help to provide the specific aspects of the data that you are interested in. For example, there is a syllabification tool that automatically syllabifies the data. Other tools can be used to reduce the different syllable tokens to abstract C(onsonant) V(owel) types. This is all extremely useful for the proposed research on syllabic development. Similarly, *Phon* can analyze the data automatically into their Place of Articulation components, providing the Place of Articulation patterns ready for analysis.

Paula Fikkert (Nijmegen University) and I will contribute our database, the CLPF database, to *Phon*. This database contains developmental data from 12 children acquiring Dutch as their first language. These children varied between 1;1.14 and 1;11.8 at the outset of the data-collecting period. Data were collected every other week, for a – depending on the child – 7 to 15 month period. The database contains about 20,000 transcribed utterances, accompanied by their audio-files. The CLPF database has until recently been available through CHILDES, but because it was a translation from a Macintosh-specific database, with a Macintosh-specific commercial font the data could not be searched very well, and could not be read on a PC.

It should be clear from the above that I wholly support this enterprise, and that I am interested in participating in the consortium organizing this new phonological database.

Yours sincerely,

Prof. Clara Levelt  
Dept. of Linguistics / ULCL  
Leiden University  
PO Box 9515  
2300 RA Leiden  
The Netherlands

e-mail: [c.c.levelt@let.leidenuniv.nl](mailto:c.c.levelt@let.leidenuniv.nl)

#### References:

- Boersma, P. & C. Levelt, (2000). A Gradual Constraint Ranking Algorithm for the Acquisition of Syllable Types, In: Clark, E. *Proceedings of the 30th Stanford Child Language Research Forum*, CSLI, 229-237.
- Fikkert, P. & C. Levelt (submitted). How does Place fall into place. Ms. Nijmegen University & Leiden University.
- Fikkert, P., C. Levelt & J. van de Weijer (2003). The Influence of Child Directed Speech on the Nature of Consonant Harmony. Paper presented at *GALA*, 4-6 September 2003, Utrecht.
- Fikkert, P., C. Levelt, S. Wauquier, M. Freitas & J. Grijzenhout, (2004). Syllabic markedness, segmental markedness, rhythm and acquisition. Paper presented at the GLOW workshop, 19 April 2004, Thessaloniki, Greece.

- Levelt, C., (1994). *On the acquisition of Place*. HIL dissertations in linguistics 8. The Hague: HAG.
- Levelt, C., N. Schiller & W. Levelt, (1999). A Developmental Grammar for Syllable Structure in the Production of Child Language. *Brain and Language* 68, 291-299.
- Levelt, C., N. Schiller & W. Levelt, (2000). The Acquisition of Syllable Types, *Language Acquisition* 8.3, 237-264.
- Levelt, C. & R. van de Vijver, (2004). Syllable Types in Cross-Linguistic- and Developmental Grammars. In: Kager, R., J. Pater & W. Zonneveld (eds.) *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge University Press, 204-218.

## 11. Lleo, Conxita

Hamburg, 15.12.2004

Dear Yvan Rose,  
Dear Brian MacWhinney,

First of all, I would like to formally express my interest in participating in the consortium organizing the new phonological database, that you are planing. During many years we have been collecting acquisition data of monolingual German and Spanish children, and lately we are collecting bilingual German-Spanish data, and we have arrived to the conclusion that unfortunately there is no single useable data base format for us, because our main concern are phonetic and phonological analyses, and the available tools up to now generally do not contemplate phonetics and phonology in an efficient way. Only recently we have access to the system EXMARaLDA, developed by our computer expert at the Collaborative Center for Research on Multilingualism, Thomas Schmidt, who is taking into consideration our phonetic needs when designing this new data base format. We have the following data from previous projects:

### PAIDUS

(spontaneous) speech, production - unstructured play sessions, 'monolingual setting' - of 5 German children in Hamburg and 4 Spanish children growing in Madrid  
monolingual children , from the onset of word production until ca. age 3 longitudinal  
analysed areas: phonetics (acoustic analyses), phonology, (proto)morphology; length of utterances (syllables and words), language use / preference

### E3 data

(spontaneous) speech, production - unstructured play sessions, 'monolingual setting'  
four German-Spanish bilingual (and a few monolingual) children , from the onset of word production until age 3-5 - longitudinal  
analyzed areas: phonetics (acoustic analyses), phonology, (proto)morphology; length of utterances (syllables and words), language use / preference

### *Data collection and preparation:*

- separate audio and video recordings, ca. 30 min.
- transcriptions: context, glosses of children's utterances, phonetic transcription of children's utterances

### *PAIDUs / E3 requirements:*

- phonetic fonts
- easy search within all data(bases) - segments, words, codings
- easy storage / copy / back up possibilities
- easy export into excel (via txt)
- compatibility with Mac and other platforms

- multi user mode

*Already available in EXMARaLDA:*

- word lists, utterance lists
- search functions
- phonetic transcriptions with unicode fonts
- syllable parser
- import from / export to PRAAT
- compatibility with Windows / Mac / Linux
- multi user

Some of the research questions mentioned above are essential in order to understand the normal development of the phonological component in German and Spanish and can serve as a basis to address further questions from theoretical, clinical and/or educational perspectives.

We would like to contribute the monolingual corpora to the database relatively soon. The bilingual data, since they constitute the main corpus of the still running project E3 of the Collaborative Center for Research on Multilingualism, being presently intensively analyzed, would have to wait a few more years (probably 2008) to be ready for being contributed.

I will be glad to answer any further questions you might have.

Best regards, Conxita Lleó

## 12. Cruz-Ferreira, Madalena

### LETTER IN SUPPORT OF THE PHON PROJECT/DATABASE

The Phon Project constitutes a much-needed breakthrough in the analysis and sharing of child language data.

My main research area within child language is multilingual phonetics, phonology and prosody/intonation, specifically during the first year of life. The use of sound in early speech provides invaluable insight into our understanding of language development, not only phonological development but also morphological, syntactic, lexical and pragmatic development. The reason is that sound is the medium of speech, and it is therefore through accurate observation of sound productions that we can make sense of speech.

Despite the recent surge of interest in very early (multilingual) speech, researchers have up to now struggled with the lack of reliable analytical tools that may enable generalisation of findings across languages and across developmental stages. This gap can now be filled through the implementation of the Phon Project.

I am currently involved in a Special Interest Group (SIG) on Child Language in Singapore, of which I am a founder and main convener. The SIG had its first meeting in September 2005. One of the issues arising from that meeting is the matter of accent, particularly English accent, given the multilingual and diglossic situation of English speakers in Singapore. For example, consonant-cluster simplification is a well-known process identified in child speech, whereby children may produce e.g. [mak] for the target /mask/, 'mask'. Singaporean children are likely to produce [mat] for the same target word, with an apparent additional substitution of [t] for final /k/. This child form is in fact a rendition of the Singaporean target /mas/, with no consonant cluster reduction, and with stopping of homorganic /s/ instead. The research questions that the Phon database will help address are:

- What are natural child forms across language varieties and across languages?
- In what ways can atypical child productions be remedied?

This issue relates to the broader issue of (un)intelligibility across language varieties, which is central to the planning of school language curricula, for example in multilingual countries where English is spoken as a non-native language, a second language or a lingua franca. The research questions that the Phon database can help address are:

- What is it that makes an English accent (un)intelligible to other users of the same language?
- In what ways can phonological/prosodic unintelligibility be remedied?

I am also involved in the creation of a Master of Arts Programme in Speech and Language Pathology, with the Faculty of Medicine and the Faculty of Arts of the National University of Singapore, for which I have been invited to teach the phonetics and phonology modules. Issues of SLP are compounded in multilingual countries, in that there often is a very thin line between speech impairment and the features of normal adult or child speech that characterise linguistic uses in these countries. The research questions that the Phon database will help address are:

### 13. Nunes, Aubrey



high quality, affordable, computer-based toys combining traditional values with good writing, design, and performance for children from 4 and up

Registered in England: No 4694778

**Pigeon Post Box Ltd, 52 Bonham Road, London, SW2 5HG**

**T: 0207 652 1347 E: [Direct@PigeonPostBox.co.uk](mailto:Direct@PigeonPostBox.co.uk)**

**I: [www.PigeonPostBox.co.uk](http://www.PigeonPostBox.co.uk)**

Director: Aubrey Nunes, PhD, FRSA, RMRCSLT

Brian MacWhinney  
CMU - Psychology  
5000 Forbes Ave.  
Pittsburgh, PA 15213  
USA

18.01.2004

Dear Brian and Yvan,

My work as a paediatric speech and language therapist evolved into a PhD at the University of Durham, England, under the title: *The Price of a Perfect System: Learnability and the Distribution of Errors in the Speech of Children Learning English*

*as a First Language.*

Essentially I relate

- a) what I would call the commonsensical observation of paediatric speech and language therapy – the typical complexity of speech and language disorders;
- b) the data I would like to offer to this project and some clinical discoveries I made myself, going beyond everyday observation;
- c) asymmetric implications between different sorts of error, as the basis for well-defined Piagetian stages;
- d) current phonological theory.

The explanation turns out to be quite difficult. The highly asymmetric error distribution by a) and b) is not easily explained. For example, coronal harmony is not prevalent, other than in particular words, all of at least three syllables in length, with a particular set of properties linking both the target and the trigger, and with other properties in the rest of the domain. Metathesis, migration, disharmony, and various sorts of floating, including one not reported before, similarly occur in very specific domains. The asymmetry does not seem to be reducible to the effects of articulation or perception. The asymmetry has to fall out from the theory.

To get this to happen, I link the theories of markedness, feature-geometry, maximally binary branchedness, context-sensitive under-specification and learnability theory. This involves the mechanism which allows phonological acquisition to proceed. To allow most children to reliably start determining the deep suprasegmental properties of the target language by around 2;3, as is observably the case, the mechanism must have some quite abstract properties, some of which are equivalent to default expectations which children bring to the learning process. Obviously these are not truly expectations because the process is unconscious. But their effect is similar to that of expectations.

An abstract speech acquisition mechanism of this sort has implications for the teaching of literacy.

An abstract mechanism of this sort seems to be independently needed in order to account for the acquisition of syntax. If so some parallelisms would be expected between syntactic and phonological acquisition. There is gross clinical evidence of this in the typical complexity of speech and language disorders and the frequency of mixed ones. There should also be reflections of this in language acquisition generally. But in order to make any such statement, it is necessary to find suitable ways of generalising across the language acquisition patterns of different individuals. This represents no small theoretical problem.

The corpus I would like to contribute consists of almost 3,000 errors in the speech of almost 100 children saying almost 100 words in a modified phonetic transcription - modified because of a) interactions between the font and the database, and b) Evelyn Abberton's idea that the phonetic transcription of incompetent speech may need (on occasion) not to be overly precise.

What I should most value would be the possibility of encoding in records of children's speech:

- a) 'sub-phonemic' distinctions such as voiceless, non-aspirates in the onset of 'spy' – effectively half-way between canonical 'pie' and 'buy';
- b) vaguely stated Abberton-type approximations. In my work, I used pointed brackets for this (rather than Abberton's circle), and within the brackets two phonetic characters separated by a slash to define the postulated degree of phonetic freedom;
- c) supra-segmental properties such as relative degrees of stress or the apparent food-structure.

I am interested in participating in the consortium organising the PhonBank database and Phon software in order to make it possible to ask questions about asymmetric implications in the patterning of naturalistically observed speech and language, and in this way to relate phonological and syntactic issues.

A first step, it seems to me, is to establish the notational conventions, such as those regarding the interpretation of the not very international IPA. I listed what I would VALUE. But I am well aware of the possibility of conflict or tension between computer and human readability.

My affiliation is commercial, but research-oriented in the framework of the company listed above.

Yours sincerely,

Aubrey Nunes

A handwritten signature in black ink, reading "Aubrey Nunes". The signature is written in a cursive, flowing style with a large, sweeping flourish at the end.

## 14. Ota, Mits

I am writing this letter in support of the completion of PhonBank project you are running with Yvan Rose and the associated development of the Phon program designed by Yvan Rose. The development of these projects will dramatically increase the range of accessible data and analytical tools that can be applied to my research projects. Most notably, I will benefit from the database and program in addressing the following sets of questions I have been investigating in my work:

### 1. The prosodic structure of early child words

To what extent is early prosodic structure composed of similar phonological constituents we find in adult language? Are the initial instantiations of such prosodic units universal such that crosslinguistic evidence points to robust general patterns (e.g., a language-independent foot structure)? Or, alternatively, do the earliest words produced by children already reflect the prosodic properties of the target language such that children exposed to typologically different languages exhibit diverse characteristics in their early word production? If so, are such language-specific child forms the products of structural restrictions or statistical properties of the input?

### 2. The development of prosody

How does the language-specific phonological system behind prosodic phenomena (e.g., stress, tone, intonation) develop? Do early production data reveal the nature of representations of these phenomena (e.g., autosegmental tones behind intonation contours) or general learning mechanisms (e.g., Operating Principles)? Are such representations and/or learning mechanism specific to individual prosodic features or are some of them general enough to apply to several domains?

It is very important that these questions be explored using crosslinguistic data from a variety of languages, as we can only tease apart common aspects and language-specific aspects of the development of prosodic structure and phenomena by comparing typologically different languages. It is equally important that the research community have a standardized data storage system and a common computational analytical program so that data and analyses can be shared and compared efficiently and reliably.

For the particular types of analysis I wish to carry out with a program such as Phon, the following features will be very useful:

1. The ability to segment audiofiles into chunks of analytical units and get them linked to transcribed data.
2. The ability to tag different levels of phonological boundaries.
3. (If possible) the ability to plug in Praat.

For the PhonBank project, I would like to contribute a spontaneous speech corpus of monolingual/non-impaired child Japanese, which I collected for my PhD dissertation (Ota 1999/2003). The corpus consists of approximately 70 hours of recordings made with three children between 1;0-2;0, 1;4-2;0, and 1;5-2;6. At this point, only the child's utterances are

transcribed, which are done in IPA and entered into an Excel file. Although a selection of utterances have been digitized, most remain on analogue audio- and super 8 videotapes.

I can also contribute data from an English/Japanese bilingual child (1;7 to 2;7), collected through double sessions (one each with a Japanese interlocutor and an English interlocutor). Only the first 9x2 (in both languages) files are transcribed in IPA, and the remaining 11x2 files are transcribed in a mixture of orthography and IPA (for 'xxx's and 'yyy's).

I look forward to the completion of the project, and I am interested in participating in the consortium organizing the Phon database.

Sincerely,

A handwritten signature in black ink, appearing to read 'Mitsuhiro Ota', with a long horizontal stroke extending to the right.

Mitsuhiro Ota  
School of Philosophy, Psychology and Language Sciences  
University of  
Edinburgh  
mits@ling.ed.ac.uk

## 15. Pater, Joe

October 13, 2004

Dr. Brian Macwhinney,  
Carnegie Mellon University  
Dr. Yvan Rose,  
Memorial University

Dear Brian and Yvan,

I am writing to formally express my interest in participating in the consortium organizing the creation of the new Phon database for phonological acquisition. I am also writing to briefly explain the research projects that I will be able to pursue once this database is ready.

One of my recent research interests has been the typology of child language processes. In Optimality Theory (Prince and Smolensky 1993), phonological processes result from rankings, or prioritizations, of constraints. Given a set of phonological constraints, it is predicted that every ranking of the constraints (an ordering from highest to lowest priority) should result in a possible language. This approach to typology has been very successful in the cross-linguistic study of phonological systems; significant advances have been made in capturing the range of possible human sound systems. I have applied this approach to child phonology in two domains: onset cluster reduction (Pater and Barlow 2003, *Journal of Child Language*), and consonant harmony (Pater and Werle 2003, *Canadian Journal of Linguistics*).

In the Pater and Barlow paper, we examine the predictions made for cluster reduction made by rankings of constraints motivated by other child language processes (and in adult phonology). In the data available to us, we found evidence of many of the predicted patterns, but a systematic check of the accuracy of the predictions was impossible, given the size of the data set. The large number of corpora that will be made available in Phon, along with the ability to conduct automated searches for instances of cluster reduction, will allow the construction of a database of cluster reduction patterns, against which our predictions, and the predictions of other theories, can be tested. Having a theory that accounts for the range of possible types of cluster reduction will certainly be a step forward in our understanding of this process, which is extremely common in phonological delay.

In studying consonant harmony, we have aimed to construct a theory that accounts for the asymmetries observed in child English. For example, assimilation of coronals to dorsals (e.g. “gog” for *dog*) is much more common than assimilation of coronals to labials (e.g. “pop” for *top*). However, our colleagues Paula Fikkert and Claartje Levelt have observed that in child Dutch, the pattern is reversed: coronal-to-labial assimilation is more common. Yvan Rose’s work on child French also points to some systematic differences between French and both Dutch and English. To better understand the sources of these

differences, we need to analyze the child corpora in exactly the same way, and also conduct analyses of corpora of child-directed speech to see if there are differences in the distribution of relevant sound sequences across languages. Both of these sets of analyses will be much easier to accomplish with the searches permitted by the Phon tools for analysis. Consonant harmony is also common in phonological delay, and understanding these cross-linguistic differences will help to pinpoint its source.

Another research project that I plan to pursue relates to the study of the relationship between phonological perception and production. Research that I have done on this topic includes Pater, Stager and Werker (2004, *Language*), and Pater (2004, in *Constraints in Phonological Acquisition*, CUP). Macken has argued in a series of papers (e.g. 1995, in *Handbook of Phonology*, Blackwell) that child processes that stop applying across-the-board are production processes, while processes that change word-by-word are the result of phonological encoding, or perception difficulties. In addition, it has sometimes been suggested that chain shifts involve misperception. The implication of these claims are two-fold: (1) across children, the same processes should change either across-the-board or by lexical diffusion, (2) neutralizations that are undone through lexical diffusion, or are involved in chain shifts, should affect contrasts that are independently known to be difficult to perceive. As far as I know, these predictions have never been examined, most likely because of the scarcity of publicly available longitudinal data. The Phon project will remove this barrier to conducting this research, and will also allow automatic identification of some of the relevant processes. A better understanding the etiology of different processes would have clear clinical implications.

I plan to contribute a corpus of phonetically transcribed child English data. These data were originally collected in a study by A.J. Compton in the 1970s, in which mothers who were speech pathologists trained in child speech transcription recorded the utterances of their children from the onset of meaningful speech. There are four children included in the corpus, which I entered into Excel spreadsheets several years ago. The size of the corpora vary, but the most detailed one has over 13,000 utterances.

Please let me know how I can be of assistance in realizing this exciting project.

Sincerely,

Joe Pater

## 16. Prieto, Pilar

October 6, 2005

**Project:** Phonetic transcription of the Serra-Solé Catalan data in CHILDES

The database for this study comprises a series of longitudinal recordings of four monolingual Catalan children, Gisela, Guillem, Laura and Pep which is part of the M. SERRA-R. SOLÉ longitudinal corpus on Catalan acquisition which can be found in the CHILDES site. Each child was video-taped on a monthly basis from the earliest productions (at about 1;0) up until four years of age. Data was collected following a naturalistic design, that is, spontaneous situations were recorded at home with the mother and the researcher.

The goal of our project is to phonetically transcribe the utterances produced by the children using IPA. We have received an undergraduate student grant for the year 2005 to undertake this project. For that, PHON provides an excellent and handy tool to transcribe using IPA symbols. Another important goal of the project is to provide links between the transcribed fragments to the actual sound/video files. In this way the researcher can go back to the actual sound and analyze prosodic features such as intonation contours, stress, phrasing strategies, etc, which is a central goal of our project. Given all this, we are really interested in participating in the consortium for organizing and setting the international phonological database coordinated by Brian MacWhinney and Yvan Rose.

The following are our specific research goals once we have the transcription of the files:

- (a) The general research goal focuses on improving our understanding of Catalan phonological acquisition, which has not been systematically studied yet.
- (b) The first specific purpose of the study is to describe the acquisition of the **segmental** material and **segmental** phonological processes (spirantization, obstruent voicing assimilation, etc.). Catalan is a language which has a rich array of phonological processes and offers a great opportunity to look at their rate of acquisition and how the interaction with other modules (namely, morphology) are being mastered by the children. Similarly, another purpose of the study is to analyze the often-mentioned phenomenon of consonant harmony. In running these analyses, the analytic features PHON proposes will be very important, as it will be crucial to identify segments in syllable-final position (for voicing for example) and in syllable-initial position (for spirantization).
- (c) The second purpose of the study is to analyze the acquisition of word-level prosodic structure, namely, syllable structure and prosodic word acquisition. We have started analyzing the acquisition of these structures in Prieto (2004) and Prieto &

Bosch-Baliarda (2004).<sup>1</sup> Our basic findings have been the following: (a) patterns of truncation in Catalan are different from neighboring languages and are guided both by maximality constraints and by frequency effects of the typical patterns found in the languages; (b) coda consonants make their appearance first in prominent positions and in stressed syllables; (c) coda consonants are not. The last two results lend support to the idea that the acquisition of phonological information is a bottom-up process that is highly constrained by prosody. In other words, prosody is the main construct which facilitates and guides the acquisition of syllabic complexity, not morphology.

- (d) The third purpose of the study is to analyze the acquisition of the **suprasegmental** processes (stress, intonation, phrasing). We know from different studies in the field that prosody acts as a mediator for speech segmentation and acquisition, but very few studies have been performed focussing on the actual production task of the actual acquisition and mastering of these features by children. Some reports in the literature have mentioned that in initial stages all syllables are stressed equally by the child and that little by little the child learns the specific stress patterns and specific irregularities of the language. We would like to get some experimental support for these observations. In running these analyses, it would be specially useful for us that PHON could provide some tools for parsing phrases, phonological phrases, and prosodic domains in general.

Yours sincerely

Pilar Prieto                      Teresa Cabra

Miquel Serra

PRIETO, P. (submitted). "Early Prosodic Word Acquisition in Catalan", *Language and Speech* (special issue on the Acquisition of the Prosodic Word, ed. by Katherine Demuth), 2004.

PRIETO, P. & M. BOSCH-BALIARDA (submitted): "The development of codas in Catalan". *Catalan Journal of Linguistics* 5, special issue on L1 Acquisition of Romance, ed. by Anna Gavarró and Conxita Lleó, 2004.

---

## **17. Ratner, Nan**

Brian MacWhinney  
CMU Psychology  
5000 Forbes Avenue  
Pittsburgh, PA 15213

Dear Professor MacWhinney:

I am greatly excited by the prospect of enhanced support for the Phon project and the enlargement of a phonological archive within Childes. As you know, I have held a long term interest in the acoustical and phonological properties of various speech registers, particularly motherese and spoken output from children with developmental disabilities such as phonological disorder, expressive language disorder and stuttering. It has been difficult to use the advantages of extensive data sets or archives in investigating these questions because of limitations on the coding and analytical programs available – they simply haven't kept pace with the advances in databased analysis of grammar and the lexicon, for example.

My most immediate need for this project will arise as part of a project being submitted for NIH funding to examine perceptual and input precursors of children's later language development. Specifically, I will be examining the frequency of phonological rule usage that blurs words boundaries or distorts canonical forms in maternal speech to young infants. The proposed utilities will greatly facilitate the coding and computational treatment of such data, which are not easily accommodated by existing programs.

There are other ongoing projects in my labs to which advances in this area could be applied. Because typically developing children, children with SLI and children who stutter share overlapping profiles of phonological ability, I would want to apply developments to my own databases of children at stuttering onset and the data from children with SLI that I have been analyzing in conjunction with Leslie Rescorla for many years. Because I am sure that a number of researchers will select proposed enhancements of coding of segmental features of speech as their primary focus, I would like to suggest that both phonological and fluency coding will be greatly improved by the current and projected abilities of Phon to handle suprasegmental aspects of speech and speech production errors well. Finally, as an Advisory Board member for the Templin data preservation project, I would be most interested in applying developments to this historically important set of archival records.

It goes without saying that I would welcome the opportunity to participate as a consortium member in this very important project. Please let me know what I can do to support this initiative further.

Cordially,

Nan Bernstein Ratner, Professor

## 18. Schneider, Katrin

October 8, 2005

Dear Yvan Rose and Brian MacWhinney,

We are very pleased to learn about your attempt to build a new database for child language, and of course it is a pleasure for us to help you extend the grant proposal with further research ideas. The aim of our project is to develop a model for the first-language acquisition of the correlates of stress in German. To approach this we analyse the effects of stress on its main acoustic correlates, i.e. on fundamental frequency, duration and the quality of different vowels in German. We aim to evaluate when children are able to perceive contrastive stress, when they are able to produce contrastive stress, and how they realize it in German.

We hypothesize that children adopt the acoustic correlates that their parents use to indicate stress and produce stress in the same way their parents do. This corresponds to the statement of Pierrehumbert ("Phonetic Diversity, Statistical Learning, and Acquisition of Phonology", *Language and Speech*, 46 (2-3), 2003, p. 127) that "[...] in early language acquisition, perception leads production." So based on the proposed database we are interested in determining the effects of stress on its main acoustic correlates, i.e. fundamental frequency, duration and the quality of different vowels, for children and their parents in German.

The results of our analyses may provide further evidence for the theories presuming that prosody is an important cue for children to split up speech input into smaller units, i.e., that prosody is essential for language acquisition. They may also serve as a basis for developing models for the acquisition of stress for different languages and, in the clinical perspective, as a basis for early diagnosis of pathological speech in children. As soon as our project is finished we will contribute our corpora to the database in question. Three corpora are recorded. The first one is a corpus of babbling of different children, the second one is a corpus of 4 to 6 children performing a TAKI task (see explanation below) with a given set of target words, and the third one is a corpus including recordings from older children (age range from 6 to 13) telling picture stories. For the second corpus we use the TAKI-task design proposed by Allen ("Development of prosodic phonology in children's speech: Further evidence from the TAKI task", in: Dressler, Pfeiffer, Rennison (eds.), *Phonologica* 1980, pp. 9-14). According to the TAKI-task pattern we created four object pairs. The objects are animal toys with bi- or tri-syllabic names. The names contain only vowels and consonants that are acquired as the first ones by German children, i.e. the vowels /a/, /i/ and /o/ and the consonants /b/, /d/, /m/ and /n/. The names within a pair of animals differ only in the position of word stress (e.g. /'bimo/ vs. /bi'mo/).

With respect to child language data, there is a need for basic acoustic analysis methods, in both the time and frequency domains. These can be separated into online analyses, which should be available at any time and at any point in the signal, continuously, and without further setup or delay; and semi-offline procedures involving the selection of

signal parts and running (non-continuous) analysis thereof. The former comprise spectrograms and F0 and formant tracking, the latter spectra of segments, with the ability to execute various spectral analysis methods like FFT or LPC analysis and with a facility to save and compare two or more separate analysis runs. Moreover, it is desirable to be able to "plug in" analysis algorithms, be it in order to use alternative algorithms for standard analyses or to integrate new or experimental methods or algorithms within the system. Thus, a facility to display analysis data in a generic way (like time/frequency/amplitude data or frequency/amplitude data, on a linear or logarithmic scale) without regard to the underlying algorithms or the units used for display. In general, analytical facilities should be designed to be easy and flexible to configure. A user should be able to extend them with the least effort possible, that is, by means of scripts or configurable analysis commands rather than extensive programming. We, the staff members of the project "Acquisition of Prosody", funded by the German Research Council (DFG) and carried out at the Institute of Natural Language Processing at the University of Stuttgart, are interested in participating in the consortium organizing this new phonological database.

Best regards,

Britta Lintfert, Katrin Schneider, Manuel Kountz and Bernd Möbius

## **19. Smit, Ann**

October 10, 2004

I am writing to indicate my enthusiastic support for the “Phon Project.” I am currently working to catalog and make accessible a very large longitudinal data set collected by Mildred Templin in the 1960s and 1970s but never published. The 436 participants were selected for their articulation skills just before they entered kindergarten. Dr. Templin did numerous assessments of articulation, language, cognition, reading, spelling, motor and perceptual processes, and personality of children who varied in articulation from pre-kindergarten through grade 11. It is my intention to make this huge data set available to all researchers by transferring it to the CHILDES site after all possible variables have been added. I have received a year of funding (2004) from the National Institute of Deafness and Communication Disorders to prepare the digital form of the existing variables.

Among the features of the Templin longitudinal project are audiorecorded conversations with the children in kindergarten and with 394 of the children in grade 11. These audiotapes have now been digitized along with some noise reduction procedures because the tapes were not made under ideal circumstances. When I turn the data set over to CHILDES, I anticipate that a number of these conversations will have already have been glossed using the CLAN procedures. However, I am interested in doing phonetic transcription using the glossed samples, and I fairly certain that other researchers might also want to do this. Therefore a robust phonetic transcription system would be very helpful. Initially, the capacity to determine percentage of consonants correct and percentage of vowels correct would be useful, as well as the ability to obtain rates of “correct” use of specific phonemes, such as /r/. A standard way to denote pitch and intensity and prosodic variation in general would also be helpful.

Because this is not a typical research endeavor, it is not possible to state exactly what kinds of research questions researchers will address. It is clear, however, that there are many possible questions, including the following:

- What aspects of future academic achievement are predictable from the child’s phonological and linguistic skills in kindergarten? (Information is available from children who have overall poor articulation and from children with specific sound errors.)
- What role do articulation and phonology play in models of brain function modeled from the database?
- How does the child’s articulation and phonological abilities affect his/her spelling?

In summary, the Phon Project looks like one that will greatly benefit the future users of the Templin data set. I would like to participate in consortium efforts to develop this project.

Ann Bosma Smit, Ph.D.  
Communication Sciences and Disorders  
asmit@humec.ksu.edu

Stemberger, Joseph

Joseph Paul Stemberger  
Department of Linguistics  
University of British Columbia

RE: ChiLDES Child Phonology Database Project

Dear Brian,

I am very interested in participating in this project. I have two sorts of databases of the speech of typically-developing children that I will be contributing to ChiLDES at some point:

- 1) extensive diary studies of two English-learning monolingual children, with a large portion of utterances in both orthography and narrow phonetic transcription; these have already been entered but not checked, and could benefit from the extended phonetic transcriptions that this project will make available; and
- 2) a cross-sectional study of monolingual Zapotec-learning children, from as young as we can get them (currently 1;11) up to about 6;0. We currently have digital video with high-quality sound for 14 children with two sessions each (ca. 80 minutes, on average), plus 4 children with a single session each (ca. 40 minutes, on average). These will eventually be transcribed in orthography and phonetics (narrowly), with English translation.

The development of standard tools to do searches and counts would be extremely useful. I would use them for basic phonological development, and also for examination of more lexical effects (including phonology-morphology and phonology-syntax interactions).

One important desideratum for these tools is that they be as theory-neutral as possible regarding representations. Not only do theories change, but there is currently a great deal of disagreement about some of the details of features/gestures, syllable structure, and foot structure. Additionally, it is always possible that children may have systems that differ from adult systems. There should be some way for the user to define features and structures, to match the user's own theoretical assumptions, and to explore the ramifications of different assumptions. In addition, frequency counts for features or segments or structures must be flexible enough that they can provide separate counts for different positions in the word, syllable, or foot, relative to boundaries of different levels. In addition, in order to allow for narrow transcriptions, the tools must support phonetic transcriptions that allow the full and extended IPA character sets, as well as any arbitrary superscripted character. Tools of this sort would facilitate and speed up research in phonological development in ways similar to the effects of other tools that ChiLDES has developed up to now.

I look forward to being involved in this project.  
Sincerely,

Joseph Paul Stemberger, Head, Department of Linguistics, UBC

## 20. Stoel-Gammon, Carol

Brian MacWhinney

CMU-Psychology

Dear Brian,

I am writing to express my support for the PhonBank project you and Yvan Rose have been working on. I currently have six doctoral students, all of whom are working on dissertations that involve phonetic transcription and phonological analysis. The dissertations focus on bilingual phonetic patterns, vowel acquisition in young children, acquisition of stops in Korean, the lexical-phonology interface in one- and two-year olds, and two studies of children with phonological disorders. My own work at this point in time involves analysis of prespeech vocalizations and of speech of children diagnosed with autism.

As you and Yvan are well aware, there is a tremendous need for analysis tools in the area of phonetic transcription and phonological analysis, much like the current CHILDES tools available for syntactic and morphological analysis. Once it is running and available to researchers, the PhonBank project will be extremely useful to all of us who work in the area of phonetics and phonology. The inclusion of a variety of corpora will also be valuable, as it will allow us to compare our findings with those of others.

I am very interested in participating in the consortium that is organizing the new phonological database. Your project is highly deserving of funding from the NIH.

Sincerely, Carol Stoel-Gammon, Ph.D.

Professor

Director, Child Speech Laboratory

## 21. Van der Pas, Brigit

Prof. B. MacWhinney  
CMU - Psychology  
5000 Forbes Ave.  
Pittsburgh, PA 15213

Utrecht, 9 october 2005

Dear Prof. MacWhinney,

I would like to express my interest in participating in the consortium organising the Phon database.

My research is concerned with the interaction of segmental and prosodic structure in first language acquisition (of Dutch). The empirical domain is formed by syllable omissions, known as 'truncation' phenomena in the literature, specifically zooming in on the segmental content of the onset position in truncated child productions. By studying these patterns, insight is gained into child selection patterns, of particular target syllables on the one hand and of particular target segments on the other. This in turn sheds new light on the role of Markedness in child phonological development. The specific research question is: do the attested child productions contain what is considered to be the unmarked segmental and syllabic material of the target form?

By studying the role of Markedness in child language data, a contribution is made to evaluating fundamental notions of Optimality Theory (Prince & Smolensky 1993). In this theory it is assumed that the initial child state is characterised by unmarked structure. Through maturation marked structure is acquired gradually by exposure to the ambient language, whereby the child-learner becomes increasingly 'faithful' to target forms. My research assesses this assumption against attested child data, collected under carefully controlled conditions.

The empirical base comprises longitudinal and cross-sectional Dutch L1 data. The corpus contains data collected by elicitation of both real words and strictly controlled nonsense words. The advantage of this experimental approach is that it provides testing environments that either do not occur in the target language, or are rare in early child spontaneous speech. Moreover, both cross-sectional and longitudinal data provide a significant source of information for developmental paths within the areas of scrutiny. The corpus involves 25 mono-lingual Dutch speakers, ranging from age 1;8 to 3;1. All data are phonetically transcribed (in IPA) and recordings are available in digital format (all nonsense word renditions have already been stored separately using PRAAT). I would be happy to contribute my database to the PHON project.

In my opinion the Phon database is a significant and long-awaited contribution to the field of phonological (and phonetic) research into language acquisition (and in general!).

By making data publicly available with immediate access to multilingual sound recordings, the reliability of phonetic transcriptions will be improved considerably. The component I am most enthusiastic about is the syllabification algorithm, enabling searches for particular segmental and prosodic structures. The kind of searches I would carry out would be, for instance, find all child productions of words with a consonant cluster in the onset of a stressed syllable (as in ‘**p**rin**ce**’, ‘sur**p**rise’); or find all truncated child productions of trisyllabic words (as in ‘banana’ -> [nana]). I would also be interested in quantitative analyses, for example, out of all multisyllabic words in a given session, how many are truncated. If the compiled data resulting from such searches can be copied into a statistical package (SPSS, Excell), data analysis would be facilitated enormously.

Unfortunately I have not been able to see the trial version of Phon yet, as I am a PC user. Obviously, I am eager to see the launch of Phon in a Windows compatible version.

Best wishes,

Brigit van der Pas  
PhD student  
Utrecht Institute of Linguistics, Utrecht University  
The Netherlands

## 22. Velleman, Shelley

October 12, 2005

Brian MacWhinney  
CMU - Psychology  
5000 Forbes Ave.  
Pittsburgh, PA 15213

Dear Brian and Yvan,

My research areas include (1) normal early phonological development across languages and dialects and (2) disordered phonological development, especially in children with childhood apraxia of speech (CAS), Williams syndrome (WS), autism spectrum disorder (ASD), and Down syndrome (DS). For the first area, I have previously studied children from the Stanford child phonology database via Marilyn Vihman; we have looked at English-, French-, Japanese-, Swedish-, Finnish-, and Welsh-exposed infants. We have identified ambient-language influences from the babbling period onward, with increasing frequency and increasing impact as the child systematizes her phonology at about 50 expressive words. I am excited about the prospect of having access to additional data on these and other languages to extend our findings. I am currently in the data-collection process of a study of phonological development in children learning African American English as a first dialect, which I plan to enter into Phon as soon as this is feasible. I believe this will be the first longitudinal phonological database from this dialect group. I am looking forward to sharing it with my colleagues via Phon as well as to using Phon for analyses of their speech at the segmental and syllabic level. Later in the project, the interface with Praat will be very helpful as I confirm my findings about trochaicization of word stress and about vowel length preceding deleted final consonants acoustically. With respect to disordered development, there are far fewer databases in existence even on paper, available by any means. I have followed two young children with WS on a long-term basis (via two years' worth of monthly videotapes shared with me by psychologist Carolyn Mervis in Louisville KY) and a few other young children more intermittently (every 3-6 months) and hope to be able transfer those transcriptions to Phon, pending Carolyn's permission and funds to pay a student to do so. I also have a set of longitudinal transcripts of 4 children with DS that could be transferred to Phon. I and my colleagues here at UMass have just received funding to explore the phonological and motor speech development of children with ASD; we will incorporate transcription via Phon into that grant proposal. My current CAS transcripts are cross-sectional, and from older children (6-7 years old), so I see sharing them via Phon as a lower priority but I would certainly be willing to do so.

I am looking forward to participating in the consortium organizing this new phonological database.

Sincerely,

Shelley L. Velleman, Ph.D., CCC-SLP  
Associate Professor

**23. Vihman, Marilyn**

Professor Brian MacWhinney  
CMU - Psychology  
5000 Forbes Ave.  
Pittsburgh, PA 15213  
October 20, 2005

Dear Brian,

I am excited at the prospect of a phonology database, which would make it possible at last for the field as a whole to move beyond the rather parochial set of assumptions - closely related to the structure of English - that have tended to guide theoretical studies of phonological development.

My primary interest is in the child's construction of a first phonological system. In order to better understand that process I have been analysing, by hand, all the relatively complete phonetically transcribed data sets that I have been able to find from as many languages as possible. I am interested in establishing empirically the range of 'adaptations' made to adult word forms by children in the early stages of word production, with the goal of arriving at quantifiable data regarding the types and extent of harmonies and melodies of various kinds. It is my impression at the moment that no theory of adult phonological structure can successfully account for or 'predict' these data. Once a large number of data sets from different children, several acquiring each of a fair sample of different languages, both phylogenetically and structurally unrelated, it should be possible to begin to draw some solid conclusions regarding the nature of the interaction of perceptual and articulatory constraints, ambient language influence and individual child effects in early word production.

For my purposes the most important characteristics of the database would be:

- reasonably exhaustive data sets from children recorded longitudinally and phonetically transcribed;
- brief phonological descriptions of the adult language and phonetically transcribed 'target words' for each child word attempt.

I am not particularly interested in other aspects of preplanned analysis, as I would prefer to see the full data set in order to carry out my own analyses. My own database, as you know, includes longitudinal recordings - either weekly or bimonthly - of children learning American English (aged 9-16 mos; N = 10) or French, Japanese, Swedish or Welsh (from 9 mos. to a 50-word vocabulary: N = 5 per language group). In addition, we are currently collecting data from 12 children acquiring British English in North Wales (aged 11 months to 24 months). I would be happy to contribute all of these data to your database, if I can obtain some support with coding and with the technical procedures involved.

Marilyn Vihman  
Chair of Developmental Psychology

## 24. Wauquier-Gravelines, Sophie

Maître de Conférences en Sciences du langage  
LLING  
UFR Lettres et Langages  
Département de Sciences du Langage et FLE  
Ch de la Censive du Tertre  
B.P. 81227  
44036 Nantes cedex 3

Tel perso : 02 40 48 68 56  
Tel bureau : 02 40 14 14 36  
fax : 02 40 14 13 27

I am researcher in acquisition of phonology for impaired (SLI) and non-impaired children. My main topic concerns the cognitive aspects (ie psychological reality) of phonological representations both for adults and children.

I actually work in the University of Nantes in a young team. We recently create a new laboratory (LLING) and develop our departement of Language Sciences. We try to develop a research and formation pole ( ? = pôle de recherche et de formation) in cognitive linguistics and acquisition. I am responsible for the « cognition » axis regrouping all the acquisition projects (phonology, syntax, semantics, pragmatics).

In the beginning, I would like to underline that « Childphon » is very interesting to study phonology in general and not only for acquisition data. It permits to record and analyse of course acquisition data, but also loanwords, and all kinds of phonological variations. Moreover Childphon is noteworthy by the fact that (ici je ne suis pas très sure de la construction syntaxique ni des mots = je veux dire « est remarquable en ceci que ou parce que ») it has been created by a phonologist who include phonological constraints in the analysis criteria (which it is not the case for other data bases). Consequently it is a very useful tool for all the phonological communities.

Finally I would like to underline that this database is very easy to use. Yvan Rose made an important effort to give a very pleasant interface, to produce mini-guides that make the utilisation of the database very easy to learn and to teach to the student who can be rapidly autonomous.

Concerning phonological acquisition in particular, it is not still very developed France. Many few data have been provided for Romance languages and especially for French. The main issues of phonological acquisition that have been investigated in English (acquisition of syllabic structures, consonant harmony, truncation, reduplication) are still not very well studied for French. Detailed studies must be developed on the questions of specific « learning-paths » and developmental aspects of French by itself, and of French compared with other romance languages and other european non-romance languages. We are only very few people working on those questions in Europe and moreless ( ? encore moins nombreux) in France, the « chilphon » could be a very useful tool for french and european laboratories working on phonological acquisition permetting an homogeneous storage and analysis of data for many different languages. It would give a very precious way

to place data at disposal of all the european community working on those issues. It would permit to create a vast homogeneous european corpus which is crucially lacking (**mettre un adverbe peut -être autre = pour traduire l'expression qui manque cruellement**). It would be a basis to exchange data and develop typological studies on acquisition of phonology. And we need to develop typological works.

I personally work (with my students) on the following topics

### **1) Acquisition of syllable structures in French**

We work on the learning-paths and developmental specificities of syllables structures in French : onsets and empty onsets acquisition, internal and final coda acquisition connected with the markedness. We need a tool permetting to observe and analyse in detail the evolution of syllables structures in longitudinal data for many children Because it apperas that all the children don't develop exactly the same learning path to deal with the syllabic structure. We get actually data of two children, but we want to built a more important data base of french children to observe the variability and study more thouroughly the question of developmental specific learning-paths. We need a way to compare stages for a child and stages between children. Childphon would give those possibilities

Those studies are connected with research in the same topics that are made by european colleagues (P. Fikkert Nimègue, C. Levelt Utrecht, M-J Freitas, Lisbonne) working on Romance and german languages and Childphon would be absolutely necessary if we want to compare and exchange regularly our data to adress a typological conception on acquisition of phonology.

### **2) Truncation and reduplication in French.**

French is a language without lexical stress. Consequently, it is very difficult to apply to French the analysis made for English (which is a lexical stressed language). We work on this question and claim that the French children do not use units like metrical feet to built progressively their lexical representation but develop a syllabic strategy based on stressed syllables inside a template. Childphon could be a very interesting tool to work on this question : it gives the possibilty to count systematically the number, the quantity (stress or unstressed) and the structure (CV, V, CVC) of the preserve syllables of the truncated syllables of a baby uterrance comparing with the adult target.

It permits also to study the reduplication strategies (which are very frequent in French) used by babies to deal with the syllabic planification of long words.

### **3) Acquisition on the French liaison.**

In french, there is a morphophonological phenomenon, a sandhi called « liaison »

On lexical boundaries V# V

- i) realization of a latent consonant on the boundary provided by the first of the two words

[1] "un enfant" : "un" [E] + "enfant" [A]fA)  
is pronounced [E]nA) fA) (*a child*)

[2] "le petit ami" : "le petit" [l'p'ti] + "ami" [ami]  
est prononcé [l'p'titami] (*a little friend*)

ii) resyllabification of the consonant on the onset of the second word

[3] [E]nA)fA) is syllabified [E\nA)\fA)], [l'p'titami] is syllabified  
[l'p'ti\ta\mi]

Children produce very regular errors of liaison consecutive to a difficulty to segment speech, between 2 and 4. We challenged two possible explanations of this critical period of errors and propose a developmental conception who connect the acquisition of liaison with the acquisition of the clitic determiner. Childphon should give tools for developing researches involving morpho-phonological phenomena.

#### 4) SLI children

My main topics are also studied for SLI children with specific phonological disorders. I actually work on those questions with M-T Le Normand (Inserm) and A. Rialland (CNRS, Paris) on « consonantless children ». Our data need to be analysed in a tool such as Childphon.

In the summary, I want to support without any hesitation the growth of Childphon database. I know Yvan Rose for a long time and I saw the database quite at the beginning. Yvan made a fantastic and very important useful work for all the community of phonologists. He spent many time to explain the possibility to use the database and responded always gently to all questions concerning the use of this tool.

Consequently I hope he will get the possibility with the new phonological data base to make profitable the time he still gave to the phonological community and could develop this tool we all need. I am interested in participating in the consortium organizing this new phonological database.

## 25. Young-Scholten, Martha

School of English Literature, Language and Linguistics  
University of Newcastle  
Newcastle upon Tyne NE1 7RU  
England

14 October 2005

Dear Yvan and Brian,

Like many others, my primary interests are in investigating actual linguistic development. But unlike most others, one of my acquisition interests - the second language (L2) development of phonology by adults - has not been at all well served by CHILDES. Moreover, the existing database in L2 phonology represents studies which primarily address issues of first language influence and eventual attainment relating to age of initial exposure. Because there are very few longitudinal studies, extremely little is known about actual L2 phonological development by either children or adults. Developmental studies are needed to determine whether L1 influence and phonological universals operate in the same manner for younger and older L2 learners and ultimately to pinpoint the source(s) of adult non-attainment in a second phonology. The mere existence of a database on phonological development dedicated to exploring development similar to CHILDES would most definitely lay down foundations for a new research program in L2 phonology.

The data I would be able to contribute represent some 33 hours of individual tape-recordings of three adolescent English-speaking learners of German. The first recordings were made in July 1996, three weeks after their initial exposure to German, which coincided with their arrival in Germany. The last recordings were made after substantial exposure to German for 11 additional months, in July 1997. Recordings were made on analogue equipment, using a Sony WM-D6C and an ECM MS-907 microphone. Because it was not feasible to use a sound booth, there is environmental noise of various types and at various decibels in the recordings. So far, the data from an experiment (picture naming) conducted during sessions to determine learners' final devoicing have been transcribed and analyzed, and the results have been published (*International Journal of Bilingualism* 1/2004). All the data have been orthographically transcribed and have been under analysis with Anne Vainikka (Johns Hopkins) for morpho-syntactic development since early 1997; these data will be made available through CHILDES once a planned book is finished.

I know too little about the operation of a phonological databases to go into any detail regarding which analytic-technical features of Phon I would need. It would be an asset if training in Phon could be offered, say, in conjunction with acquisition conferences. Of course any phonological acquisition database should be able to allow (or at least to facilitate) analysis of development at the segmental and supra-segmental levels. With respect to my own data, I am interested in being able to determine the development of certain German vowels, of /r/ and of primary stress, and I imagine this is what would also be of interest to others using Phon for the analysis of L2 German. Finally, it would undoubtedly be an advantage if one could analyze phonological data in conjunction with morphological data, given the recent proposals regarding morphology-phonology interfaces (e.g. Goad's work).

For all the above reasons, I am very interested in participating in this project.

Dr. Martha Young-Scholten  
Senior Lecturer in Second Language Acquisition

## 26. Zamuner, Tania

Brian MacWhinney  
CMU - Psychology  
5000 Forbes Ave.  
Pittsburgh, PA 15213

October, 2005

Dear Yvan Rose & Brian MacWhinney :

I am interested in participating in the consortium to organize Phon, a new phonological database. Below is a description of the project on which I will be working with the phonological database, and a description of the database that I will eventually submit to CHILDES. This is part of a project entitled "The Nature of Representations in Developmental Speech Perception and Production," awarded to myself (Tania S. Zamuner), from the Netherlands Organisation for Scientific Research (NWO). This longitudinal project is in collaboration with Paula Fikkert at the University of Nijmegen.

I am also planning to use Phon as a tool for analyzing experiments looking at children's phonological acquisition. For example, I am currently testing children's productions of morpho-phonological alternations in Dutch. Children between 30-32 and 42-44 months of age are being testing on how their spontaneous and imitative productions of singular and plural Dutch nouns. Phon will be used to analyze children's responses. Children's responses will be linked to the audio files of their productions and this will allow for blind coding of experimental sessions. The program can then be used to search for particular word types to look at production patterns across a larger group of children.

Sincerely,

Tania Zamuner

University of British Columbia  
Department of Psychology  
2136 West Mall Vancouver,  
BC V6T 1Z4  
Canada  
e-mail: TZamuner@psych.ubc.ca  
phone: (604) 822-6408  
Fax: (604) 822-6923

Project Description

Background

Arguments in favor of detailed or abstract representations in the acquisition of phonology are partly divided by the fields of speech perception and production, respectively. This

difference is mirrored in language acquisition research. Studies of infant speech perception have shown that infants have acquired detailed knowledge of their language's sound patterns. In studies of children's speech production, it is argued that children's early words do not reflect this early perceptual knowledge, but rather that the sound patterns seen in children's words reflect abstract representations. This project is conducting more detailed analyses of children's perception and production abilities compared to children's lexicons, and looking at perception and production in tandem.

### Approach

Particularly relevant to Phon, a component of this project will look at the developing lexicon and its relationship to perception and production skills. The goal is to track the development of children's lexicons and how this relates to their ability to perceive and produce language. Subjects will be studied from 18 to 28 months of age. Parents will be asked to track the lexical development of their children by completing a diary and a standardized questionnaire used in the field. Children will be tested once a month on how they perceive and produce specific sounds and lexical items, using tasks similar to those in the first set of studies. Analyses will compare children's lexical development to the development of perception and production skills.

### Innovation

What makes this project unique is that it addresses this question by bridging the gap between our understanding of both perceptual and production abilities. This project is also important because it focuses on a specific period in development which has not yet been examined. Previous studies have not mapped the development or linked the development of children's perceptual and production abilities to children's lexicons. Additional advantages of this project include *i)* the possibility to gain more insight into the relationship between children's sound awareness and their later reading ability and impairments, (since research on typically developing children establishes baselines to which children with language impairment can be compared), *ii)* a detailed study of the acquisition of Dutch will enrich our understanding of language development given that most conclusions are based on the acquisition of English alone, and *iii)* the longitudinal perception and production data will be made publicly available on the CHILDES database, where perception and production data are currently not available, yet crucial for analyses and theories of acquisition.

## 27. Zmarich, Claudio



to Prof. Brian MacWhinney  
CMU – Psychology  
5000 Forbes Ave.  
Pittsburgh, PA 15213

### Sezione di Padova **Fonetica e Dialettologia**

Via Anghinoni 10 35121 Padova  
Tel 049 8274418 Fax 049 8274416  
e-mail: [pagnin@pd.istc.cnr.it](mailto:pagnin@pd.istc.cnr.it)  
http: [www.pd.istc.cnr.it](http://www.pd.istc.cnr.it)

In this letter I would like to express my feelings of appreciation for your great and long-lasting efforts in creating the largest database of children's speech, currently available to the scientific community, and the necessary specialized tools for drawing the best from it (CHILDES).

I strongly hope that this unique work will continue, and will be aimed at matching the needs of those researchers mainly interested in the phonetic and phonological aspects of children's speech, like me. In fact, CHILDES mostly deals with the morphological and syntactic aspects of language development, while no suitable tools for phonology are currently available.

A basic and very general research question I would like to address based on the proposed database and specialized tool like PHON, concerns the emergence of the Italian phonology. A substantial agreement exists, among researchers, over the fact that the phones and types of syllables produced at the beginning of babbling are drawn from a restricted set of potential phones and types of syllables constituting languages, and that their frequency in children is universal and not influenced by characteristics specific to the native language (Vihman, 1996). But children finally come to speak a particular language, so it is of utmost importance to detect, describe and explain this process of differentiation. In order to do so, it is necessary to demonstrate, at least, that: (1) the phonetic differences among national groups are greater than differences within each national group; (2) these differences reflect the phonetic patterns peculiar to any national language. This kind of comparison could be managed only by means of data sharing and broad-based investigations.

Another interesting question, to me, is represented by the primacy of the jaw in modeling the segmental intra- and inter-syllabic structure of babbling, as put forward by P. MacNeilage and B. Davis, versus the primacy of the active articulators which do not include the jaw and which causes the inventory of the earliest segments to be based on differences in places of articulation, emerging very early from a natural partition of the vocal organs, as put forward by L. Goldstein. The automatic calculation of C-V co-occurrences allowed by PHON on the one hand, and of the segmental matches between children productions and attempted targets on the other, will make the elaboration process much faster and more error-free.

This kind of research is important from a theoretical perspective, but it also maintains a clinical value, because the description of the phonetic development of the typical children will be the norm to which the atypical children will be compared.

I would like to contribute to the database with a corpus constituted by the phonetic transcription of one normally developing Italian child, biweekly audio-recorded from 6 to 18 months and then monthly audio-recorded from 19 to 27 months. However, depending on financial support and/or on students/transcribers, I could include further nine children. Six of them were recorded from 6 to 30 months and three from 8 to 30 months. Some of the recordings are already phonetically transcribed.

The organization of this new database is of great interest to me and I hope I will be able to contribute to it.

With kindest regards,

Padova, 13-10-2005  
Dr. Claudio Zmarich