

Laboratoire de Linguistique Formelle

www.llf.cnrs.fr

1. Presentation
2. Research
3. Valorisation
4. Phd Students



1. Presentation

- University- Paris Diderot & CNRS Research laboratory created in 1972
- Director: A. Abeillé, vice-director C. Plancq
- Members:
 - 14 CNRS researchers
 - 17 University Professors
 - 2 engineers + 1 administrator
 - 32 PhD students
 - 6 postdocs and lecturers
- 2013: national five-year evaluation : A+

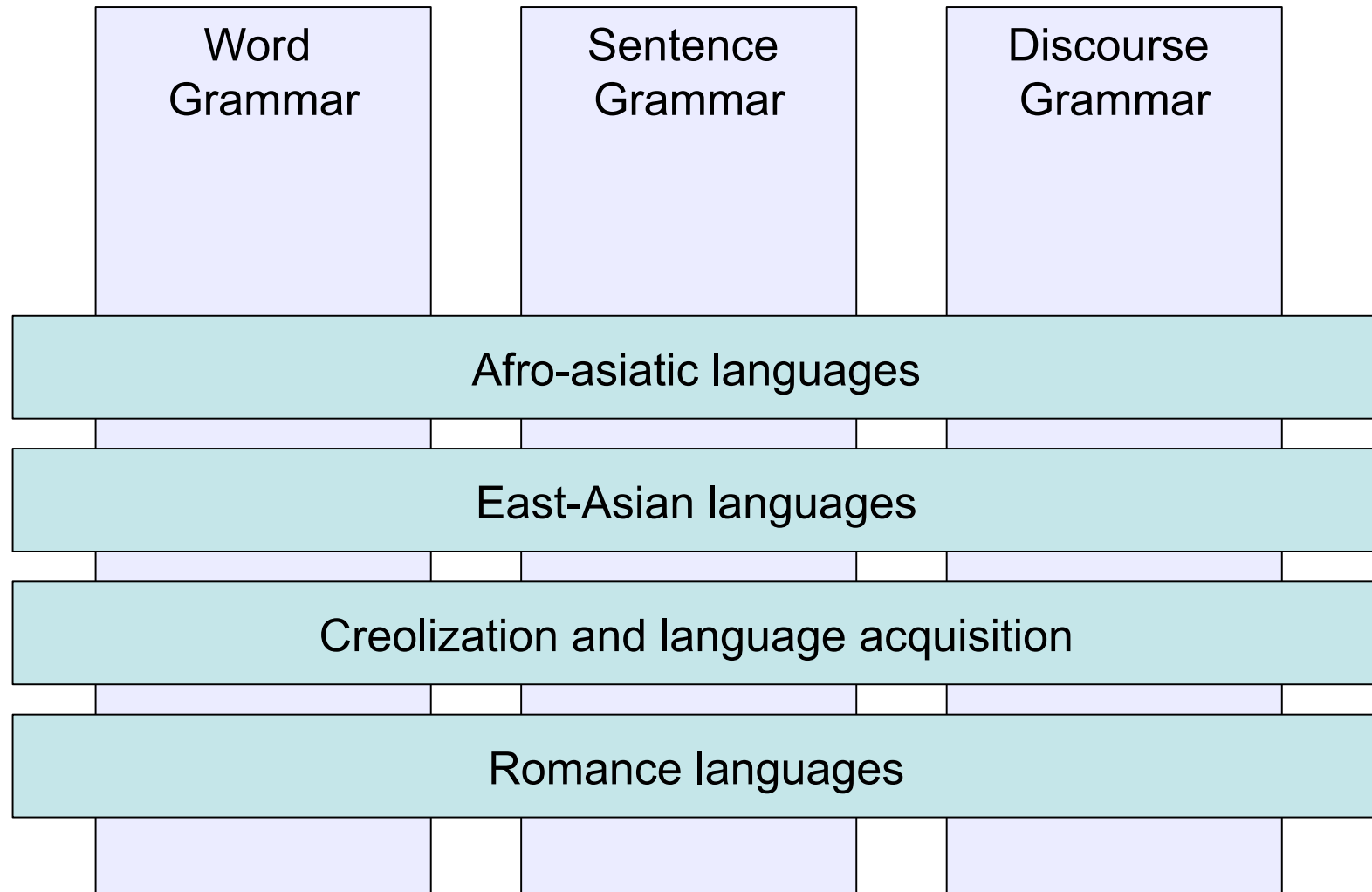


2. Research

- General perspective: the study of language through the diversity of languages
- Strong points:
 - Specialists of most aspects of language (phonology, morphology, syntax, semantics, discourse) and their interfaces
 - Specialists of most modern linguistic theories (DM, PFM, minimalism, HPSG, DRT, SDRT...)
 - Diversity of languages studied (more than 70)
 - Variety of competences (formal linguistics, computational linguistics, sociolinguistics, psycholinguistics, phonetics...)
 - Variety of resources (library, corpora...) and experimental material (eye-tracker)



Scientific strands



Scientific production

- **2007-2012**: 120 journal articles, 30 books, 170 book chapters, 107 invited conferences, 157 proceedings publications
- **Linguistic Resources** (on-line or distributed for research purposes):
 - Annotated corpora : French treebank (FTB), spoken corpora for French (ACSYNT, ANR Pro-gram, ANR Etape), physiological corpus (with eye-movement measures), multilingual corpora (Chinese, Japanese, Amis, Taroko ANR Clapoty), ancient texts in Mauritian creole ...
 - Lexical DataBases: French determiners across time (ANR Elico), inflected lexica for French, Kurdish, Persian, Portuguese..., lexica for derived French words (Converts, Dénalex...)
 - Computer grammars for NLP (German, Hausa...)
 - First-hand data (fieldwork) on some endangered languages (Australia, Oman...)



Scientific collaborations

- **LabEx** *Empirical Foundations of Linguistics* (2011-2021)
10 year project with 13 Parisian labs (LPP, LLF, LIPN/RCLN, Llacan, Alpage, Lattice, CRLAO, MII, HTL, LACITO, SEDYL, LPPCog, Memcog)
LLF in charge of 2 strands and 20 work packages
- **3 CNRS projects** (PEPS *French/German machine translation*, TUL: *Denumerals across languages*, *Identifying events*)
- **3 ANR projects** : Phorevox, Asfalda, Oman-sam
- **2 European projects** : 7th PRCD Tameal, ESF (COST) Parseme
- **Other international projects**: Capes Cofecub (Florianopolis), ESF Networks (Pisa), Hubert Curien (Coimbra), GDR-i (*Creole languages*)...



3. Valorisation

Reference grammar books

Coordination of *Grande Grammaire du français*, *Reference Grammar of Romanian*, contribution to *Chinese Reference Grammar* and to several international handbooks

Scientific Edition

Brill's Annual of Afroasiatic Languages and Linguistics, *Oxford Bibliography in Linguistics*...

Commercial licences sold for the Frenchtreebank (Google, Microsoft, Apple, Sony...)

Contracts with DGLFLF, Agence universitaire de la francophonie (AUF), OIF...

Other activities

Fête de la science, Expolangues, Babel 93...

Newspapers and radios (Libération, RFI, France culture...)

Invitations in French representations abroad

Expertises for museums (musée du Quai Branly, Maison de la culture yiddish...)



4. PhD Program

- Reorganization of Doctoral Schools and Masters' Programmes inside the PRES *Sorbonne Paris Cité*
 - LLF strongly involved in the Linguistics Doctoral School project and in the *Linguistique théorique, descriptive et expérimentale* Masters' programme



PhD Program

2008-2013: 25 PhD theses and 5 habilitations defended at LLF

Most of them have found academic positions in CNRS, French Universities (Pau, Paris 3, Inalco, Rennes 2, Lille 3, Tours) or abroad (Academy of Science in Bucarest and Budapest, Tokyo U, Nat Seoul U, U Jerusalem, Singapore etc)

2013: 35 PhD students registered (21 with financial support from Ecole Doctorale, LabEx, MAE, International funding...)

- Involved in LLF research, access to funding for material and travels
- Regular seminars and intensive training
- Summer schools (Experimental Grammar june 2013)
- Invited professors: in 2013 : D. Büring (Vienne), J. Fodor (CUNY), E. Williams (Princeton), G.T. Stump (Kentucky), I. Sekarina (Staten Island), J. Runner (Rochester), J. Bresnan (Stanford), S. Lappin (King's College)



Research axis:

Languages of East Asia

- **Permanent members** : 5 (D. Paillard, M.-C. Paris, C. Saillard, V. Pan, P. Martin)
- **Doctoral students** : 15 (in 2012), including 9 on Chinese
- **International collaborations** (Beijing University, Academia Sinica, Hong Kong Polytechnic...)



Languages studied

- Sinitic languages : Mandarin Chinese
- Altaic language family : Korean, Japanese
- Austro-asiatic family : Khmer, Vietnamese, Thai
- Austronesian family : Amis, Taroko



Languages of East Asia: recent facts (1)

- 5 professors/researchers supervising PhD dissertations in the field
- 10 PhD dissertations finished during the last 5 years (among which 7 about Chinese)
- 14 ongoing PhD dissertation (among which 7 about Chinese)
- 2 habilitation theses since 2010



Languages of East Asia: recent facts (2)

- Chinese Linguistics: organization of an international conference (EACL VII, 2011) and a Spring School (2011)
- (South-)East Asian Linguistic : international conference and scientific meeting (2010);
- Publications (by permanent members) : 1 book, 23 publications, more than 60 oral communications in the last 5 years



Languages of East Asia: recent research topics

- Plurality, numeral approximation and reduplication (Chinese, Khmer)
- Coordination (Chinese & Vietnamese)
- Particles (Chinese, Vietnamese, Khmer)
- Interrogatives & quantification (Chinese)
- TAM markers (Chinese)
- Etc...



Languages of East Asia: current research topics

- Conjunction & comitativity in East Asian Languages (resp. M.-C. Paris, D. Paillard)
- Interrogatives & exclamatives in Chinese (resp. V. Pan)
- Discourse markers in East Asian Languages (resp. D. Paillard, C.Saillard)



Relevant work packages in LabEx EFL

- Dative constructions in Chinese (resp. C.Saillard) & annotation of Chinese Corpora (collaboration with Alpage)
- Comparison of discourse markers in Birman, Khmer, Thai and Vietnamese (resp. D. Paillard)
- Multifactorial Analysis of Language Change (coordination Sedyl)



Research on Chinese

Presentation of ongoing
research by lab members

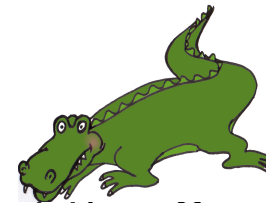
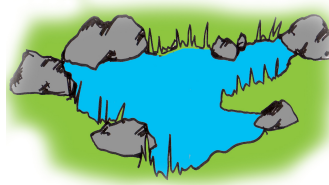
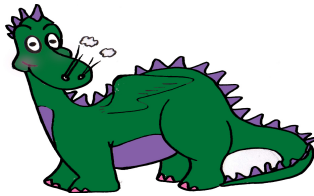


The acquisition of French object clitics by Mandarin Chinese learners of French (B. Hemforth)

- Factors that play a role in the acquisition of a language are
 - The age at which the language is acquired
 - Cognitive capacities of the learner
 - For late learners, their native language
- French object clitics are particularly hard to learn for children as well as for learners of French as a second language



Production experiment



Le dragon et le crocodile sont tous les deux près de la rivière. Cependant, le crocodile veut aller de l'autre côté pour voir ses autres amis. Du coup, le dragon boit toute l'eau de la rivière afin que son ami le crocodile puisse passer. Il y a beaucoup d'eau dans la rivière car il a beaucoup plu ces derniers temps.

- Qu'est-ce que fait le dragon avec l'eau de la rivière?
- Le dragon..... **le** boit. (The dragon it drinks.)
- The correct answer in Chinese would correspond to «The dragon drinks.» without the object pronoun.



Results

- French speakers produce the correct object clitic in about 83 percent of the cases. **17 % errors**
- Chinese (Mandarin) speakers produce the object clitic in about 28 percent of the cases. **82% errors**
- The higher the working memory capacity of the Chinese (Mandarin) speaker, the more correct object clitics he/she produces.
- Even speakers who started learning French late were much better when their working memory capacity was high.



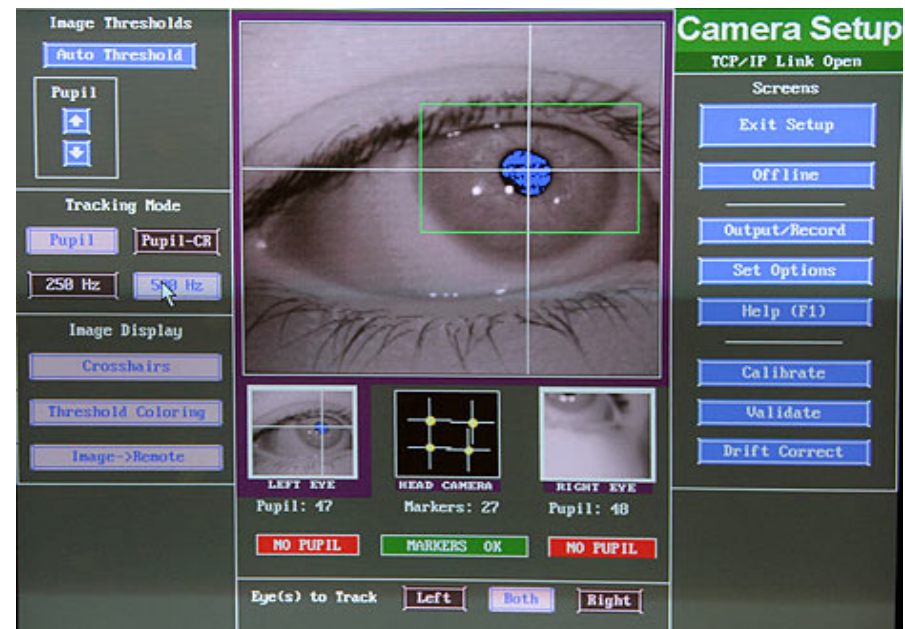
Experiments on French and Chinese (Mandarin) relative clauses

- French relative clauses are always following the noun they describe
 - L'étudiant qui suit le cours avec attention
- Mandarin Chinese relative clauses precede the noun they describe
 - 认真地听着课的孩子
- Question: How do bilingual Chinese-French speakers understand relative clauses in their first and in their second language



Eyetracking experiments

- Participants hear French or Chinese sentences like:
- Trouvez l'escrimeur qui/que dessine la princesse.
- 请找画击剑员的公主/画公主的击剑员



Classifying Dialogue Utterances

Kwong-Cheong Wong and Jonathan Ginzburg '**Investigating Non-Sentential Utterances in a Spoken Chinese Corpus**' **Proceedings of PacLing 2013, Tokyo**

Dialogue is full of fragmentary utterances which are incomplete sentences but nevertheless convey a complete sentential meaning in the given context—Non sentential utterances.

Building on the work by
Fernández and Ginzburg on English NSUs in the British National Corpus ,
this paper describes an extension of this approach to Chinese.

The work uses the Mandarin part of the
National Chengchi University (NCCU) Corpus of Spoken Chinese,



Classifying Dialogue Utterances

(4) *Classification Ellipsis* is a reprise fragment used to clarify not fully comprehended utterance. E.g.,

A: 那個 要 測 .. 測 就 好 了
na4 ge yao4 shan4 shan4 jiu4 hao3 le
that CL have-to boil boil then OK PRT
'That one has to be boiled. Just boil it.'

B: 測 嘅
shan4 o
boil PRT
'Boil?'

[5: 197-198]

(12) *Verbal Affirmative Answer* is an affirmative response which contains just a modal verb. E.g.,

A: 那個 真的 會 睡著 嗎
na4ge zhen1de hui4 shui4zhao2 ma
that really will sleep:RESULT QST
'Does it really make people sleepy?'

B: 會
hui4
right
'Right.'

[6: 109-110]



The DUEL Project

- Joint with Bielefeld University under the French-German bilateral programme of ANR and DFG (under review)
- Aims: to through an empirical program comparative French, German, and Chinese data on the frequent,
- but under-studied spoken language phenomena dysfluency, exclamations, and laughter.
- The project will involve building parallel corpora in French, German, Chinese
- The project will develop a detailed grammar for the parallel Chinese-French-German corpus of dysfluencies, incl. Work on editing phrases like Chinese 'en' 嗯, 'nage' 那个.
- The project will compare the repertory of laughter uses across French, German, Chinese and develop a semantic/pragmatic analysis of these.
- The formal grammatical work will inform a computational implementation in a spoken dialogue system.



Chinese Linguistics

Ongoing PhD research

