

Zusammenfassungen (in alphabetischer Reihenfolge der Beiträger)
 Abstracts (in alphabetical order of contributors)
 Résumés (en ordre alphabétique par auteurs)

Balthasar, Lukas / Dzian, Elodie / Heiden, Serge Outils de mise en place de corpus multimédiaux : le cas de la linguistique interactionnelle (A9.1)	1
Bazzanella, Carla / Bosco, Cristina Rethinking the <i>modal shift</i> in Old and Modern Italian corpora (B6.2)	1
Becker, Martin / Gleßgen, Martin / Kunstmann, Pierre / Matthey, Anne-Christelle / Stein, Achim Corpus électroniques d'ancien français: ressources, outils, exploitation (A1)	2
Beeching, Kate Politesse et changement linguistique : le cas des particules pragmatiques en français contemporain (B15.1)	3
Bellini, Daniele / Schneider, Stefan The <i>Banca dati dell'italiano parlato</i> (BADIP): Present functions and future perspectives (A6.2)	4
Bert, Michel / Balthasar, Lukas / Mondada, Lorenza La base de données "Corpus de langues parlées en interaction" du GRIC : Etat des lieux, perspectives d'évolution, usage dans le cadre d'études diachroniques en linguistique interactionnelle (A9.2)	4
Bick, Eckhard / Módolo, Marcelo Letters and Editorials: A grammatically annotated corpus of 19th century Brazilian Portuguese (A8.2)	5
Bourova, Viara / Slobbe, Bianca A la recherche du 'conditionnel latin': les périphrases <i>infinitif</i> + forme de <i>habere</i> examinées à partir d'un corpus électronique (B1.2)	5
Brasseur, Patrice L' <i>Atlas linguistique et ethnographique normand</i> : lemmatisation, indexation (A3.1)	6
Clavería Nadal, Gloria / Torruella Casañas, Joan Base de datos para un corpus de documentaciones léxicas (A7.1)	7
Costa Giménez, Elisabet / Guarro, Beatriu / Martínez i Taberner, Caterina / Molina Compte, Caterina / Pons, Lúdia / Sayós i Santigosa, Rosa / Tisselli, Eugenio La llengua catalana als espots de TV. Software per a l'anàlisi (A10.1)	7
Davies, Mark Advanced research on syntactic and semantic change with the 100 million word, fully-annotated <i>Corpus del Español</i> (A5.1)	8
Davies, Mark On the frequency, use, and omission of <i>se</i> : evidence from the 100 million word <i>Corpus del Español</i> (B14.1)	9
Decorte, Steven Linguistic annotation in the Integrated Language Database of 8th-21st-Century Dutch (ILD) (A4.1)	9
Dufresne, Monique / Dupuis, Fernande / Tremblay, M. La cartographie des préverbes en français médiéval (B5.2)	10
Eggert, Elmar Analyse eines Korpus zu französischen Wohnernamen (<i>gentilés</i>). Die Anwendbarkeit sprachlicher Regelmäßigkeiten in der automatischen Sprachverarbeitung (A3.2)	10
Enrique-Arias, Andrés Evolución histórica de la posición de los marcadores de objeto en español desde una perspectiva tipológico funcional: un estudio de corpus (B12.2)	11

Eufe, Rembert	
Sprachwandel in venezianischen Kolonialverwaltungsdokumenten um 1500 (B4.2)	11
Finck Ziller, Diomara	
Funcionalidade e integração: Expansão e redução de cláusulas finais na fala do sul do Brasil (B16.1)	12
Fischer, Susann	
The Case of Old Romance clitic placement revised: Data from Old Catalan (B12.1)	12
Forcada, Mikel L. / Garcia-Sempere, Marinela / Garrido, Alícia /	
Gilabert-Zarco, Patrícia / Iturraspe Bellver, Amaia / Montserrat, Sandra	
Desenvolupament a base de corpus d'un analitzador morfològic de català antic (A7.2)	14
Gallegos Shibya, Alfonso	
Nominalización y registro técnico. Algunas relaciones entre morfopragmática, tradiciones discursivas y desarrollo de la lengua en español (B2.2)	14
Gerner, Hiltrud	
La Base des Lexiques du Moyen Français (BLMF) (A2.2)	15
Guillot, Céline / Heiden, Serge / Prévost, Sophie	
Corpus historiques – balisage, étiquetage, exploitation	15
Hakulinen, Soili	
Corpus de traductions françaises du <i>De consolazione philosophiae</i> de Boèce (B3.2)	15
Hinzelin, Marc-Olivier	
Die Untersuchung der Stellung der Objektpronomina zum finiten Verb anhand eines Korpus zur (alt)okzitanischen Syntax (B10.2)	16
Jodl, Frank	
Der Einsatz kommerzieller Übersetzungs-Software bei der Erforschung von sprachlicher Innovation und Sprachwandel: Der Modus nach ‚wenn, falls‘ im Portugiesischen, Brasilianischen und Spanischen (B16.2)	17
Kabatek, Johannes	
How can we measure discourse traditions in historical corpora? (B11.1)	17
Kaiser, Georg A.	
Bibelübersetzungen als Grundlage für empirische Sprachwandeluntersuchungen (B8.1)	17
Klein-Andreu, Flora	
El papel del ‘corpus’ en una investigación histórica funcional (B14.2)	18
Klump, Andre	
Zur Herausbildung der konsonantischen Besonderheiten des Spanischen auf der Insel Hispaniola (B9.1)	19
Knutsen, Anne Moseng / Ploog, Katja	
La grammaticalisation de LA en (français) abidjanais : une “marque” à tout faire ? (B13.2)	19
Langford, Chad / O’Connor, Kathleen M.	
Micro-change in the French noun phrase: the case of pre-posed color adjectives (B13.1)	20
Lavrentiev, Alexei	
Représentation de transcriptions diplomatiques de manuscrits français médiévaux dans le système XML-TEI (Projets <i>Charrette</i> et <i>Queste du Graal</i>) (A2.1)	20
López, Marisol	
El <i>Corpus del Gallego Actual</i> (CORGA). Problemas de configuración y anotación (P)	21
Marchello-Nizia, Christiane	
Pour une linguistique historique ‘outillée’ (P)	21
Martins, Ana Maria	
Passive and impersonal <i>se</i> in the history of Portuguese (P)	21
Massip, M. Àngels / Veny, Joan	
El Corpus de Diatopia Diacrònica de la Llengua Catalana (CDDL) (A5.2)	21

Moser, Karolin	
Cortesía y descortesía en el español de Costa Rica: hacia una perspectiva diacrónica (B15.2)	22
Peemöller, Johanne	
Graphiegeschichte und Sprachwandel: Möglichkeiten und Grenzen der Untersuchung von Sprachwandel in den ältesten Zeugnissen der französischen Sprachgeschichte (B5.1) ...	22
Perissinotto, Giorgio	
The Historical Dimension of Spanish in the Western United States: Notes for a Lexical Corpus and Bilingual Glossary (A8.1)	22
Rey, Christophe	
Les sons dans l' <i>Encyclopédie Méthodique</i> (B11.2)	23
Ross, Sarah H.	
A corpus-based approach to register variation in Early Latin (B1.1)	23
Scarano, Antonietta / Signorini, Sabrina	
<i>Corpus linguistics</i> e variazione diacronica. Uno studio su <i>corpora</i> dell'italiano dagli anni '60 ad oggi (A6.1)	24
Sinner, Carsten	
Estudio longitudinal de cambios lingüísticos inducidos por contacto lingüístico (B2.1)	25
Sitaridou, Ioanna	
A corpus-based diachronic study of null subjects in French (B10.1)	25
Stark, Elisabeth	
Explaining article grammaticalization in Old Italian (B6.1)	26
Stein, Peter	
Die Übersetzungen von Livius' <i>'Ab Urbe condita'</i> in die romanischen Sprachen, 1323 bis heute. Vorstellung des Textcorpus und der Analysemöglichkeiten (A3.1)	27
Thun, Harald	
Corpora der gesprochenen Sprache und Texte der literarisierten Mündlichkeit als Zeugnisse des Sprachwandels (B7.1)	28
Videsott, Paul	
Skriptometrische Analyse Nordostitaliens: einige Fallbeispiele (B4.1)	28
Völker, Harald	
Non-lineare Hypertextstrukturen in Textkorpora (A10.2)	28
Wanner, Dieter	
Data ex machina: searching for explanations in Spanish diachrony (P)	29
Wesch, Andreas	
Textos jurídicos y administrativos españoles del primer período de la expansión colonial (B9.2)	29
Verzeichnis der Beiträge · List of contributors · Index des auteurs	31

Zusammenfassungen (in chronologischer Reihenfolge nach Sektionen)
 Abstracts (in chronological order according to panels)
 Résumés (en ordre chronologique par sections)

Becker, Martin / Gleßgen, Martin / Kunstmann, Pierre / Matthey, Anne-Christelle / Stein, Achim Corpus électroniques d'ancien français: ressources, outils, exploitation (A1.1; A1.2)	3
Lavrentiev, Alexei Représentation de transcriptions diplomatiques de manuscrits français médiévaux dans le système XML-TEI (Projets <i>Charrette</i> et <i>Queste du Graal</i>) (A2.1)	20
Gerner, Hiltrud La Base des Lexiques du Moyen Français (BLMF) (A2.2)	15
Brasseur, Patrice L' <i>Atlas linguistique et ethnographique normand</i> : lemmatisation, indexation (A3.1)	6
Eggert, Elmar Analyse eines Korpus zu französischen Wohnernamen (<i>gentilés</i>). Die Anwendbarkeit sprachlicher Regelmäßigkeiten in der automatischen Sprachverarbeitung (A3.2)	10
Decorte, Steven Linguistic annotation in the Integrated Language Database of 8th-21st-Century Dutch (ILD) (A4.1)	9
Davies, Mark Advanced research on syntactic and semantic change with the 100 million word, fully-annotated <i>Corpus del Español</i> (A5.1)	8
Massip, M. Àngels / Veny, Joan El Corpus de Diatopia Diacrònica de la Llengua Catalana (CDDL) (A5.2)	21
Scarano, Antonietta / Signorini, Sabrina <i>Corpus linguistics e variazione diacronica. Uno studio su corpora dell'italiano</i> dagli anni '60 ad oggi (A6.1)	24
Bellini, Daniele / Schneider, Stefan The <i>Banca dati dell'italiano parlato</i> (BADIP): Present functions and future perspectives (A6.2)	4
Clavería Nadal, Gloria (Barcelona) / Torruella Casañas, Joan Base de datos para un corpus de documentaciones léxicas (A7.1)	7
Forcada, Mikel L. / Garcia-Sempere, Marinela / Garrido, Alícia / Gilabert-Zarco, Patrícia / Iturraspe Bellver, Amaia / Montserrat, Sandra Desenvolupament a base de corpus d'un analitzador morfològic de català antic (A7.2)	14
Perissinotto, Giorgio The Historical Dimension of Spanish in the Western United States: Notes for a Lexical Corpus and Bilingual Glossary (A8.1)	22
Bick, Eckhard / Módolo, Marcelo Letters and Editorials: A grammatically annotated corpus of 19th century Brazilian Portuguese (A8.2)	5
Balthasar, Lukas / Dzian, Elodie / Heiden, Serge Outils de mise en place de corpus multimédiaux : le cas de la linguistique interactionnelle (A9.1)	1
Bert, Michel / Balthasar, Lukas / Mondada, Lorenza La base de données "Corpus de langues parlées en interaction" du GRIC : Etat des lieux, perspectives d'évolution, usage dans le cadre d'études diachroniques en linguistique interactionnelle (A9.2)	4
Costa Giménez, Elisabet / Guarro, Beatriu / Martínez i Taberner, Caterina / Molina Compte, Caterina / Pons, Lúdia / Sayós i Santigosa, Rosa / Tisselli, Eugenio La llengua catalana als espots de TV. Software per a l'anàlisi (A10.1)	7

Völker, Harald	
Non-lineare Hypertextstrukturen in Textkorpora (A10.2)	28
Ross, Sarah H.	
A corpus-based approach to register variation in Early Latin (B1.1)	23
Bourova, Viara / Slobbe, Bianca	
A la recherche du ‘conditionnel latin’: les périphrases <i>infinitif</i> + forme de <i>habere</i> examinées à partir d’un corpus électronique (B1.2)	5
Sinner, Carsten	
Estudio longitudinal de cambios lingüísticos inducidos por contacto lingüístico (B2.1)	25
Gallegos Shibya, Alfonso	
Nominalización y registro técnico. Algunas relaciones entre morfopragmática, tradiciones discursivas y desarrollo de la lengua en español (B2.2)	14
Stein, Peter	
Die Übersetzungen von Livius‘ ‘ <i>Ab Urbe condita</i> ’ in die romanischen Sprachen, 1323 bis heute. Vorstellung des Textcorpus und der Analysemöglichkeiten (A3.1)	27
Hakulinen, Soili	
Corpus de traductions françaises du <i>De consolatione philosophiae</i> de Boèce (B3.2)	15
Videsott, Paul	
Skriptometrische Analyse Nordostitaliens: einige Fallbeispiele (B4.1)	28
Eufe, Rembert	
Sprachwandel in venezianischen Kolonialverwaltungsdokumenten um 1500 (B4.2)	11
Peemöller, Johanne	
Graphiegeschichte und Sprachwandel: Möglichkeiten und Grenzen der Untersuchung von Sprachwandel in den ältesten Zeugnissen der französischen Sprachgeschichte (B5.1) ...	22
Dufresne, Monique / Dupuis, Fernande / Tremblay, M.	
La cartographie des préverbes en français médiéval (B5.2)	10
Stark, Elisabeth	
Explaining article grammaticalization in Old Italian (B6.1)	26
Bazzanella, Carla / Bosco, Cristina	
Rethinking the <i>modal shift</i> in Old and Modern Italian corpora (B6.2)	1
Thun, Harald	
Corpora der gesprochenen Sprache und Texte der literarisierten Mündlichkeit als Zeugnisse des Sprachwandels (B7.1)	28
Kaiser, Georg A.	
Bibelübersetzungen als Grundlage für empirische Sprachwandeluntersuchungen (B8.1)	17
Klump, Andre	
Zur Herausbildung der konsonantischen Besonderheiten des Spanischen auf der Insel Hispaniola (B9.1)	19
Wesch, Andreas	
Textos jurídicos y administrativos españoles del primer período de la expansión colonial (B9.2)	29
Sitaridou, Ioanna	
A corpus-based diachronic study of null subjects in French (B10.1)	25
Hinzelin, Marc-Olivier	
Die Untersuchung der Stellung der Objektpronomina zum finiten Verb anhand eines Korpus zur (alt)okzitanischen Syntax (B10.2)	16
Kabatek, Johannes	
How can we measure discourse traditions in historical corpora? (B11.1)	17
Rey, Christophe	
Les sons dans l’ <i>Encyclopédie Méthodique</i> (B11.2)	23
Fischer, Susann	
The Case of Old Romance clitic placement revised: Data from Old Catalan (B12.1)	12

Enrique-Arias, Andrés	
Evolución histórica de la posición de los marcadores de objeto en español desde una perspectiva tipológico funcional: un estudio de corpus (B12.2)	11
Langford, Chad / O'Connor, Kathleen M.	
Micro-change in the French noun phrase: the case of pre-posed color adjectives (B13.1)	20
Knutsen, Anne Moseng / Ploog, Katja	
La grammaticalisation de LA en (français) abidjanais : une “marque” à tout faire ? (B13.2)	19
Davies, Mark	
On the frequency, use, and omission of <i>se</i> : evidence from the 100 million word <i>Corpus del Español</i> (B14.1)	9
Klein-Andreu, Flora	
El papel del ‘corpus’ en una investigación histórica funcional (B14.2)	18
Beeching, Kate	
Politesse et changement linguistique : le cas des particules pragmatiques en français contemporain (B15.1)	3
Moser, Karolin	
Cortesía y descortesía en el español de Costa Rica: hacia una perspectiva diacrónica (B15.2)	22
Finck Ziller, Diomara	
Funcionalidade e integração: Expansão e redução de cláusulas finais na fala do sul do Brasil (B16.1)	12
Jodl, Frank	
Der Einsatz kommerzieller Übersetzungs-Software bei der Erforschung von sprachlicher Innovation und Sprachwandel: Der Modus nach ‚wenn, falls‘ im Portugiesischen, Brasilianischen und Spanischen (B16.2)	17

Balthasar, Lukas / Dzian, Elodie / Heiden, Serge (Lyon)

Outils de mise en place de corpus multimédiaux : le cas de la linguistique interactionnelle (A9.1)

La linguistique interactionnelle et ses précurseurs utilisent depuis une trentaine d'années des bandes sonores et des enregistrements audiovisuels d'interactions comme ressource empirique de leur recherche. La profondeur historique des données ainsi accumulées permet aujourd'hui des analyses diachroniques dans ce domaine. Toutefois, les conditions pour mener à bien des recherches sur ce type de matériau ne sont données que depuis peu de temps (et elles évolueront sans doute encore considérablement).

L'objectif de cet exposé consiste à présenter un état des lieux concernant les diverses technologies – logicielles en particulier – qui facilitent aujourd'hui la mise en place de corpus interactionnels de haute qualité.

Au moyen d'un exemple concret, nous présenterons l'emploi successif des *logiciels de transcription* actuellement utilisés dans les laboratoires GRIC et Anacoluthé pour la mise en place de corpus, ainsi que la fonctionnalité de divers *utilitaires* conçus par nous-mêmes. Ces derniers servent d'abord au transfert des sorties des divers logiciels entre eux ; elles servent aussi à la création d'un format « pivot » XML (SMIL) qui sera interprétable par les divers logiciels de transcription et qui facilitera l'exploitation des transcriptions dans les bases de données modernes.

Les étapes de transcription et de conversion présentées dans notre exposé sont les suivantes :

- Transcription de base des aspects verbaux d'une interaction dans le logiciel *Praat*© Boersma & Weenink
- Conversion des fichiers Praat (.TextGrid) vers .TXT (incorporation des transcription dans des éditeurs de texte comme MS Word© et impression directe sur papier ; ©SergeHeiden)
- Conversion des fichiers Praat (.TextGrid) vers .Interaction-XML (format pivot pour des transcriptions d'interactions multimodales – ce format est actuellement en stade de prototype ; ©SergeHeiden)
- Analyse prosodique dans *Praat*
- Importation / exportation des fichiers .Interaction-XML dans le logiciel *Anvil*© Kipp et transcription des aspects non verbaux des interactions (Anvil<>Interaction-XML ©SergeHeiden)
- Annotation prosodique des productions verbales et annotation des aspects non verbaux d'une interaction dans *Anvil*
- Importation/exportation des fichiers .Interaction-XML dans le logiciel *CLAN*© CHILDES PROJECT et transcription des aspects non verbaux des interactions (CLANl<>Interaction-XML ©SergeHeiden)
- Transfert des fichiers entre les divers logiciels Praat, CLAN et Anvil
- Présentation (et navigation dans) de fichiers .Interaction-XML dans les browsers habituels (MS Internet Explorer©, Netscape© etc.) ■

Bazzanella, Carla (Torino) / Bosco, Cristina (Torino)

Rethinking the *modal shift* in Old and Modern Italian corpora (B6.2)

Some discourse markers such as *allora*, *anzi/anziché*, *contra*, *or/ora*, *piuttosto che*, *poi che/poiché*, *siccome* will be considered and their diachronic development analyzed.

Generally, in Old Italian different values were co-present, while in Modern Italian only one meaning, and one given value (usually the modal one), have prevailed. For example, *anzi* has kept only its modal value, and lost its spatial and temporal values which co-occurred in Old Italian, sometimes with the same author (cf. Bazzanella forthc. a, b, c). The general trend by

which “Meanings tend to become increasingly based in the speaker’s belief state / attitude toward the proposition” (Traugott 1989, 35) seems to be confirmed.

The data on Old Italian will be taken from *Italant* corpus (i.e. *Per una grammatica dell’italiano Antico*: a corpus of texts written in Florentine between 1250 and 1300; cf. Renzi and Bisetto 2000, see *Gatto*; <<http://www.csovi.fi.cnr.it/gframe.htm>>), and compared with the data on Modern Italian. With regard to the data on Modern Italian, several available Italian corpora will be referred to, among which is Turin University Treebank corpus (cf. Bosco *et al.* 2000, Bosco 2001 <<http://www.di.unito.it/~tutreeb>>).

Anzi and *contra* will be analyzed also in reference to their Latin correspondents *ante* / *contra* (cf., respectively, Bazzanella *forthc.*, Orlandini *forthc.*).

On the basis of the data resulting from the corpora (which, as we shall see, provide the analyst with a useful tool with regard both to completeness and easy retrieval), the shift “from propositional through textual to expressive” (cf. Traugott 1982, 256) will be investigated, and the “complexity of ‘grammaticalization chains’” (cf. Brinton 1996, 65) will be argued for.

Bibliographical references

- Bazzanella, Carla *forthcoming a*: Dal latino *ante* all’italiano *anzi*: la ‘deriva modale’; in Garcea, Alessandro (ed.): *Colloquia Absentium. Studi sulla comunicazione epistolare in Cicerone*. Torino: Rosenberg.
- Bazzanella, Carla *forthcoming b*: Discourse Markers and Politeness in Old Italian; in Held, Gudrun (ed.): *Partikeln und Höflichkeit*. Wien: Peter Lang.
- Bazzanella, Carla *forthcoming c*: Discourse Markers in Italian: towards a ‘compositional’ meaning; in Fischer, Kerstin (ed.): *Approaches to discourse particles*. Berlin/New York: De Gruyter.
- Bosco, Cristina / Lombardo, Vincenzo / Vassallo, Daniela / Lesmo, Leonardo 2000: Building a treebank for Italian: a data-driven annotation schema; in Gavrilidou, Maria et al. (eds.): *Proceedings of LREC2000*, Athens, Greece, 29-30 May 2000. Athens: National Technical University of Athens Press, 99-105.
- Bosco, Cristina 2001: Grammatical relations system in treebank annotation; in Miltsakaki, Eleni et al. (eds.): *Proceedings of Student Research Workshop of Joint ACL/EACL Meeting*, Toulouse, 1-6.
- Brinton, Laurel J. 1996: *Pragmatic markers in English: grammaticalization and discourse functions*. Berlin/New York: Mouton de Gruyter.
- Orlandini, Anna. *forthcoming*: *Contra*; in Garcea, Alessandro (ed.): *Colloquia Absentium. Studi sulla comunicazione epistolare in Cicerone*. Torino: Rosenberg.
- Renzi, Lorenzo / Bisetto, Antonietta. (eds.) 2000: *Linguistica e italiano antico*; numero monografico di *Lingua e stile* 35.4, 537-743.
- Traugott, Elizabeth C. 1982: From propositional to textual and expressive meanings: some semantic-pragmatic aspects of grammaticalization; in Lehman Wilfred. P. / Malkiel Y (eds.): *Perspectives on Historical Linguistics*, Amsterdam: Benjamins, 245-271.
- Traugott, Elizabeth C. 1989: On the rise of epistemic meanings in English: an example of subjectification in semantic change; *Language* 57, 33-65. ■

Becker, Martin (Stuttgart) / Gleßgen, Martin (Zürich) / Kunstmann, Pierre (Ottawa) / Matthey, Anne-Christelle (Zürich) / Stein, Achim (Stuttgart)

Corpus électroniques d’ancien français: ressources, outils, exploitation (A1)

1. (Martin-D. Gleßgen, Zürich / Achim Stein, Stuttgart) En partant de la présentation de certaines ressources électroniques de l’ancien français, notamment du corpus d’Amsterdam (Dees / van Reenen) et d’un échantillon lorrain des “Plus anciens documents linguistiques de la France”, nous montrerons dans quelle mesure elle pourront être utilisées pour le traitement automatique de l’ancien français. Nous discuterons les résultats de l’analyse morphologique et de la lemmatisation obtenues grâce à la combinaison du “TreeTagger”, un étiqueteur automatique qui a été entraîné sur le corpus d’Amsterdam et enrichi d’autres ressources lexicales, et du lemmatiseur philologique “Phoenix”; nous parlerons brièvement des possibilités d’annotation syntaxique.

2. (Pierre Kunstmann, Ottawa : « D’un corpus à l’autre : Précisions sur le changement morphologique et syntaxique en ancien français ») Pour illustrer l’exploitation du corpus d’Amsterdam dans le cadre du changement linguistique, nous nous proposons de suivre, brièvement, en diachronie plusieurs points intéressant de la morphologie et la syntaxe de l’ancien français (mor-

phèmes indéfinis, constructions relatives, interrogatives, emploi des temps et des modes après certaines conjonctions). Les données du corpus d'Amsterdam seront comparées, d'une part, avec celles d'un corpus plus restreint: les 5 romans de Chrétien de Troyes, minicorpus établi au LFA, synchronique certes (seconde moitié du 12e s.), mais avec une certaine variation interne); d'autre part, avec celles d'un corpus plus vaste en aval: la Base Textuelle de Moyen Français de l'ATILF.

3. (Martin Becker, Stuttgart) Afin d'illustrer les possibilités d'exploitation offertes au chercheur par le Corpus d'Amsterdam, nous entendons scruter un aspect primordial dans l'évolution de la langue française, notamment la corrélation entre le système des conjonctions et les formes verbales du passé, plus précisément le passé simple et l'imparfait. Nous proposons donc une étude de cas qui, en dépouillant les données obtenues grâce à des coupes transversales, tranche une question plus globale: Les conjonctions contribuent-elles à un procès d'abstraction de la signification des catégories tempo-aspectuelles qui se soustraient de plus en plus à la "détermination locale" exercée par le sémantisme du verbe (l'aspect verbal) pour se transformer en des marqueurs d'une perspective plus complexe au service de la disposition ou de l'arrangement d'une topologie textuelle – avec une physionomie tout en saillie et en creux?

4. (Anne-Christelle Matthey, Zürich) Notre illustration se base sur un corpus des "Plus anciens documents linguistiques de la France" au sein duquel nous analyserons les chartes conservées aux A.D. de la Meuse. Ces documents administratifs (1220-1270) sont à considérer dans une logique de formation des langues écrites; il s'agira ici d'étudier leur degré de normalisation, en confrontant la stéréotypie des chartes à leurs variances graphématique et morphologique. Nous présenterons plus en détail les fonctionnalités du lemmatiseur "Phoenix" utilisé pour cette interrogation. ■

Beeching, Kate (Bristol)

Politesse et changement linguistique : le cas des particules pragmatiques en français contemporain (B15.1)

Les études diachroniques de la langue française, se basant sur l'examen minutieux de documents écrits, ont traditionnellement misées sur des facteurs internes linguistiques pour en expliquer l'évolution. Croft (2000: 4) constate pourtant que le changement se produit au sein de la situation de communication orale dans la bouche des locuteurs. Qui plus est, quelques études plus récentes nous attirent l'attention au rôle de la routinisation et au lien entre variation synchronique et diachronique (Heine / Claudi / Hünnemeyer, 1991) et à celle des inférences invitées généralisées (GIINS, Traugott / Dasher, 2002). L'avènement des corpus oraux – et surtout leur naissante longévité – nous permet de discerner des changements en cours.

Cette communication vise à donner un résumé préliminaire des résultats d'une comparaison des fréquences distributionnelles de certaines particules pragmatiques dans le Corpus Orléans (1966-1970) et le Corpus Bristol (1980-1990). L'étude de ces corpus sociolinguistiques conforte l'assertion de Jakobson (1963 [1952]: 37) selon qui 'un changement est ... à ses débuts, un fait synchronique.' La théorie de la politesse (Eelen, 2001), qui souligne l'importance primordiale de la sociabilité, explique l'ubiquité de certaines expressions qui médient les rapports humains, telles les formes généralement qualifiées de 'tics' ou de 'remplisseurs'. La présence massive de ces formes presque désémantisées, leur polysémie, leur routinisation et leur tendance vers l'emblématisation (identification avec une classe, un groupement ou une identité social) les rendent particulièrement susceptibles au changement. ■

Bellini, Daniele (Trieste) / Schneider, Stefan (Graz)

The *Banca dati dell'italiano parlato* (BADIP): Present functions and future perspectives (A6.2)

Since February 2003 linguists working in the field of spoken language analysis and corpus linguistics have at their disposal a web-based database (cf. <<http://languageserver.uni-graz.at/badip>>) containing an enriched edition of the 500.000 word LIP corpus, the largest and most important corpus of contemporary spoken Italian. Access to all functions of BADIP is free. The database is part of the Language Server of the University of Graz (Austria), a web platform that provides information and materials on languages, language learning and language analysis. At the moment, BADIP is being supported by a fund of the Austrian Ministry for Education, Science, and Culture.

Already in its present state BADIP is a highly developed instrument. Besides viewing the texts of the LIP corpus in several formats and downloading them, it allows several types of researches. As well as a number of basic querying functions (single word, multiple word, word sequences, wild cards for characters and words), it makes possible researches based on annotations. The great majority of the corpus is enriched with POS-tags and lemmata, other annotations contain information on overlaps, paralinguistic phenomena, intonation and so forth. More data are being added continuously. Users may save the syntax of their own queries as well as the results and export them.

In the future BADIP will develop in several directions. Major interventions will concern the enhancement of the query functions, the inclusion of other data on the LIP corpus, as well as the adaptations necessary for another corpus of spoken Italian to be placed on the platform. ■

Bert, Michel / Balthasar, Lukas / Mondada, Lorenza (Lyon)

La base de données "Corpus de langues parlées en interaction" du GRIC : Etat des lieux, perspectives d'évolution, usage dans le cadre d'études diachroniques en linguistique interactionnelle (A9.2)

La base *Corpus de langues parlées en interaction* (CLAPI) a fait l'objet d'une première présentation lors de la 1. *Freiburger Arbeitstagung zur Romanischen Korpuslinguistik*.¹ L'enrichissement et l'exploitation de cette base constituent l'une des activités du GRIC (Groupe de Recherche sur les Interactions Communicatives), son développement bénéficie d'un soutien important et son évolution au cours des derniers douze mois a donc été considérable, ouvrant des perspectives nouvelles. Les objectifs de notre intervention sont les suivants :

- Présentation de l'état actuel de la base CLAPI : caractéristiques technologiques de la base de données (structure informatique de la base et de sa mise en ligne, nature et fonctions des descripteurs, protocole d'utilisation et gestion des droits d'accès par le web, etc.), types et caractéristiques des données hébergées (les formats audio et vidéo, les formats des transcriptions), etc.
- Esquisse des développements futurs de CLAPI : conception d'outils de mise en place et d'analyse de corpus assistées par ordinateur, mise en ligne de corpus interactionnel sous la forme de documents multimédias hautement structurés et *streamables* (documents XML / SMIL), permettant l'interrogation conjointe de transcriptions avancées et d'enregistrements en temps réel, etc.
- Démonstration de l'usage de la base de données au moyen d'exemples concrets, comme l'évolution de la nature et de la récurrence des chevauchements des productions verbales

¹ Cf. Traverso & Bruxelles (2002), *Les corpus de langue parlée en interaction au GRIC*, dans Pusch & Raible (éds), *Romanische Korpuslinguistik, Romance Corpus Linguistics*, Tübingen : Gunter Narr

dans le corpus “Physique”. Ce corpus comporte des données collectées au cours des quinze dernières années dans des cours de physique en lycée ; les transcriptions de ces données permettront à terme des observations relativement fines concernant l’évolution des comportements interactionnels entre les élèves, ainsi qu’entre ces derniers et leurs enseignants.

Cette présentation permettra de mettre en évidence le fonctionnement et l’intérêt scientifique d’une base de données interactionnelles technologiquement très développée. Dans son état actuel, CLAPI présente déjà une dimension longitudinale importante, et son enrichissement favorisera l’ouverture de nouveaux champs de recherche, tels que celui d’une linguistique interactionnelle diachronique. ■

Bick, Eckhard (Viby) / Módolo, Marcelo (São Paulo)

Letters and Editorials: A grammatically annotated corpus of 19th century Brazilian Portuguese (A8.2)

This paper describes the structural design, philological editing and grammatical annotation of a 175.000 word corpus of 19th century Brazilian Portuguese (PB), consisting of newspaper leaders and letters to the editor (Barbosa & Lopes 2002). The corpus was compiled as part of a planned diachronic *megacorpus*, within the *Projeto Para a História do Português Brasileiro*, a joint effort of researchers in eight Brazilian states, and has already been used in ongoing efforts to describe the social history of PB, as well as the evolution of grammatical changes in the language. The material was grammatically annotated with the PALAVRAS parser (Bick 2000), a robust Constraint Grammar system using word based tags for word class, lemma, inflexion and syntactic function, as well as directional dependency markers at both group and clause level. Historical spelling variations were treated by systematic orthographical filtering, lexicon additions and morphological heuristics. Preliminary test runs on small parts of the corpus indicate error rates of 2.5% for word class and 7–8% for syntactic function. Though the system performs twice as well on current texts, these numbers suggest that a “modern” CG-parser can achieve results on 19th century data with only minor alterations. Especially in connection with manual revision, an annotated version of the corpus is expected to facilitate quantitative research, structural pattern searches and category based queries.

References:

- Barbosa, Afrânio & Lopes, Célia (eds), *Críticas, queixumes e bajulações na Imprensa Brasileira do séc. XIX: cartas de leitores e cartas de redatores*. Rio de Janeiro: UFRJ, 2002
- Bick, Eckhard, *The Parsing System ‘Palavras’ – Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus: Aarhus University Press, 2000. ■

Bourova, Viara (Antwerpen) / Slobbe, Bianca (Antwerpen)

A la recherche du ‘conditionnel latin’: les périphrases *infinitif* + forme de *habere* examinées à partir d’un corpus électronique (B1.2)

On se propose de communiquer les résultats actuels du projet de recherche de l’Université d’Anvers “Le développement du conditionnel dans les langues romanes (français, italien, roumain)” qui envisage, à partir d’une analyse sémantique et syntaxique des occurrences, et en tenant compte de la répartition géographique et chronologique, d’expliquer l’apparition de cette nouvelle forme dans le système verbal roman.

Le conditionnel, qui n’existait pas en latin, provient, d’après l’hypothèse classique, d’une périphrase du type *cantare habebam / habui*. Jusqu’à l’heure actuelle, il n’a pas été l’objet d’une étude diachronique approfondie: la démarche la plus fréquente est d’évoquer son développement

comme analogique à celui du futur (Fleischman 1982). D'autre part, les études consacrées aux périphrases avec infinitif + forme de *habere* n'ont pas beaucoup élargi leur base après l'article fondamental de Thielmann (1885): les plus récents (Coleman 1971, Pinkster 1987) s'appuient sur les exemples du Thesaurus Linguae Latinae qui pour l'essentiel reprend ceux de Thielmann.

L'apport majeur de notre approche est de partir d'un corpus beaucoup plus vaste, la CLCLT-5 (on ne tient compte que des œuvres antérieures au 9e siècle, soit 26 millions de mots environ) et complétée par d'autres bases électroniques consultables sur le web. L'extraction des périphrases avec infinitif + forme de *habere* a permis la constitution d'une propre base de données sous Access dont le résumé statistique et l'analyse syntaxique et sémantique permettront de:

1. vérifier les hypothèses traditionnelles et de voir dans quelle mesure l'émergence du conditionnel peut être liée à celle du futur
2. se prononcer à partir d'un large ensemble de faits attestés sur la pertinence de l'hypothèse générativiste de D'Hulst (2001) qui, partant du cadre de Giorgi & Pianesi (1997), postule un développement en trois phases, en expliquant les transitions d'une phase à l'autre par le déplacement d'information temporelle à des projections fonctionnelles plus hautes.
3. présenter un bilan provisoire des avantages mais également des restrictions d'une recherche diachronique sur corpus électroniques.

Bibliographie :

- CLCLT - 5: *Library of Latin Texts (CD-rom)* release 2002, moderante Paul Tombeur, Centre Traditio Litterarum Occidentium. Turnhout, Brepols Publ.
- Coleman, R. (1971) 'The origin and development of Latin *habeo* + infinitive.' In : *Classical Quarterly* 21, pp. 215-232.
- D'Hulst, Yves (2001) 'French and Italian Conditionals: From Etymology to Representation.' To appear in: Guéron & Lecarme (eds.) *The Syntax of Tense and Aspect. Proceedings of the International Conference on the Expression of Tense and Aspect, Paris December 2000*. MIT.
- Fleischman, Suzanne (1982) *The future in thought and language: diachronic evidence from Romance*, Cambridge (Cambridge studies in linguistics; 36)
- Giorgi, Alessandra & Pianesi, Fabio (1997) *Tense and Aspect. From Semantics to Morphosyntax*. New York – Oxford, Oxford University Press
- Pinkster, Harm (1987) 'The Strategy and Chronology of the Development of Future and Perfect Tense Auxiliaries in Latin.' In: Harris & Ramat (eds.) *Historical Development of Auxiliaries*. Mouton de Gruyter, Berlin – New York – Amsterdam, pp. 193-223.
- Thesaurus Linguae Latinae (1925-1934), vol.VI:3, Leipzig, Teubner
- Thielmann, Ph. (1885) 'Habere mit dem Infinitiv und die Entstehung des romanischen Futurums.' In : Wolfgang, E. (ed.) *Archiv für die lateinische Lexicographie und Grammatik mit Einschluss des älteren Mittellateins*. Leipzig. ■

Brasseur, Patrice (Avignon)

L'Atlas linguistique et ethnographique normand : lemmatisation, indexation (A3.1)

L'Atlas linguistique et ethnographique normand (ALN) est en voie d'achèvement. Trois volumes ont été publiés (1980, 1984, 1997). Le dernier est en cours de préparation. Sa publication marquera la fin d'un processus commencé en 1970, qui a nécessité 7 années d'enquêtes dans 114 points répartis dans les 5 départements normands et les Iles anglo-normandes, à partir d'un questionnaire de plus de 2200 questions.

On a parfois reproché aux atlas linguistiques régionaux français d'être difficilement comparables entre eux, puisque les intitulés des cartes ne se recouvrent que très partiellement. Une recherche informatique par forme phonétique pose des problèmes de tous ordres et n'est pas non plus envisageable à partir des données publiées. L'utilisation de ces ouvrages passe plutôt par l'indexation de l'ensemble des formes et, au préalable, par leur lemmatisation. Ce travail, qui a été entrepris par plusieurs auteurs, est maintenant achevée pour les 3 premiers volumes de l'ALN.

Cette communication se propose de décrire la procédure de lemmatisation des formes à partir des notations phonétiques assorties des références (cartes et listes de mots). Outre l'accès à des attestations éparses dans plusieurs cartes, et même dans des volumes différents, le regroupement des formes ainsi opéré peut mener à de nouvelles hypothèses étymologiques. Mais l'une des utilisations majeures d'une telle base de données, comprenant un total provisoire de 27 500 entrées, est sans doute, à terme, la rédaction d'un dictionnaire des parlers normands comportant un aspect encyclopédique issu des données ethnographiques. Nous en donnerons un premier exemple, en rédigeant et commentant l'article *rouge*. ■

Clavería Nadal, Gloria (Barcelona) / Torruella Casañas, Joan (Barcelona)

Base de datos para un corpus de documentaciones léxicas (A7.1)

La presente comunicación expone el trabajo que se está realizando en el Seminario de Filología e Informática de la Universitat Autònoma de Barcelona y que consiste en la realización de un corpus léxico a partir de las documentaciones que aparecen en el *Diccionario Crítico Etimológico Castellano e Hispánico* de Joan Corominas y José Antonio Pascual. Este trabajo se integra dentro del proyecto de *Informatización del DCECH* que dirige la Dra. Gloria Clavería.

El diccionario aporta casi cincuenta mil documentaciones para ilustrar los más de ocho mil lemas que presenta (algunos lemas vienen avalados por sólo una documentación, pero otros pueden llegar a aportar hasta siete o más). Todas estas documentaciones van acompañadas de una serie de datos, explícitos unos y implícitos los otros, que, debidamente sistematizados y organizados, pueden ofrecer al investigador un importante banco de datos del que se puede extraer informaciones imposibles de obtener con la consulta del diccionario en su soporte original, que es de seis volúmenes en papel. Nuestro trabajo, en este apartado del proyecto, ha consistido, principalmente, en hacer explícitos aquellos datos que en la obra aparecían como implícitos (por ejemplo, autores o fechas deducibles a partir de la obra) y en reducir a un común denominador las diferentes variantes formales de los datos que aporta el *DCECH* (por ejemplo, conversión a una sola escala de las diferentes maneras de reseñar la fecha de una obra), con el fin de que se puedan manejar con el gestor de una base de datos.

El objetivo de nuestra comunicación es presentar la estructura de la base de datos que se ha diseñado para organizar todas las referencias de manera que, mediante ordenaciones y filtros, se puedan sacar el máximo de informaciones, así como mostrar una primera radiografía del corpus usado por los autores del diccionario: primeras documentaciones aportadas, periodos más representados, autores más explotados, etc. ■

Costa Giménez, Elisabet / Guarro, Beatriu / Martínez i Taberner, Caterina / Molina Compte, Caterina / Pons, Lúdia / Sayós i Santigosa, Rosa / Tisselli, Eugenio (Barcelona)

La llengua catalana als espots de TV. Software per a l'anàlisi (A10.1)

La nostra comunicació té per objecte presentar un projecte d'investigació sobre la llengua de la publicitat. Aquest projecte consisteix en la formació i explotació científica d'una base de dades informatitzada i digitalitzada integrada per anuncis de televisió. Els materials són una part dels espots comercials emesos pels canals de la televisió catalana durant la dècada compresa entre 1990 i 1991. La base de dades ha de servir per:

- L'estudi de la llengua catalana de la publicitat i especialment l'estudi dels recursos lingüístics usats en la creació d'eslògans i espots.
- L'estudi dels factors psíquics i socials que el publicista canalitza a través de la llengua a fi de persuadir o de seduir un receptor.

El contingut de la comunicació se centraria en dos aspectes: En primer lloc, explicariem el tractament dels materials que formen el corpus i, en segon lloc, presentariem la fitxa que hem creat a partir de la qual podem portar a terme la catalogació i estudi de la base de dades. ■

Davies, Mark (Provo)

Advanced research on syntactic and semantic change with the 100 million word, fully-annotated *Corpus del Español* (A5.1)

The newly-released, 100 million word *Corpus del Español* (<<http://www.corpusdelespanol.org>>) is the first publicly-available corpus to allow advanced research on diachronic Spanish syntax and semantics. Unlike the CORDE corpus from the Real Academia Española (www.rae.es), the Corpus del Español is annotated for parts-of-speech and lemma, and also allows advanced pattern-matching queries and user-defined syntactic and semantic annotation. The following are just a few examples of the limitless queries that can be performed with the corpus:

1) Parts-of-speech, lemma, and synonyms, e.g.:

[query] !mandar.* que *.v_subj_ra
 [explanation] [any form of any synonym of *mandar*] + *que* + [-ra past subjunctive]
 [results] *hizo que dijeran, mandaron que uinieran*

2) The queries can also be limited and sorted by historical period, e.g.:

[query] lo *.adj que ser.* [limits] 1900s>2 –1800s –1700s
 [explanation] *lo* + [adjective] + *que* + [any form of *ser*], which occurs more than two times in 1900s, but not in 1700s or 1800s
 [examples] *lo difícil que es, lo importante que era*

3) Users can also create and re-use (as part of the query syntax) their own customized lists (of any length) of words that are related syntactically or semantically, e.g.:

[query] me/te/nos/os/le/le [garcía:emociones].* que
 [explanation] [one of (me/te/nos/os/le/le)] + any form of any of the verbs in the [emociones] list that [garcía] has created + *que*
 [examples] *te gustaría que, me sorprendió que, nos molestaba que*

4) Semantic shift, as evidenced by changes in collocations or the frequency of synonyms, e.g.

[query] *.n suave [limits]+1800s –1700s
 [explanation] which nouns occur with *suave* in the 1800s but not the 1700s, possibly indicating a change in the meaning of *suave*
 [examples] *viento suave, acento suave*

[query] !destruir [limits]+1900s -1700s
 [explanation] synonyms of *destruir* that occur in the 1900s, but not in the 1700s
 [examples] *desintegrar, desmantelar*

In addition to the advanced queries, the corpus is also very fast. All of the preceding queries can be run against the entire 100 million word corpus in less than three seconds. ■

Davies, Mark (Provo)

On the frequency, use, and omission of *se*: evidence from the 100 million word *Corpus del Español* (B14.1)

To date, there has been no full-scale investigation into diachronic shifts in the use of *se* in Spanish. Monge (1954) is perhaps the most complete account of *se* in Old Spanish, but is limited primarily to that period. Kemmer (1993) provides a good theoretical framework, but is much weaker in actual data. Studies such as Maldonado (1999) are theoretically and empirically engaging, but focus on just a small subset of the uses of *se*. Martín Zorraquino (1979) is perhaps the most complete account to date, but is based on a relatively small corpus of texts.

This paper is an attempt to use data from the newly-completed, 100 million word *Corpus del Español* (<<http://www.corpusdelespanol.org>>) to answer several basic questions about the frequency, use, and omission of *se* in Spanish, which – surprisingly – have not yet been adequately studied to this point. These questions include:

- 1) With which verbs was there the largest increase in the use of *se* from Old Spanish (1200s-1400s) to the Spanish of the Golden Age (1500s-1600s)?
- 2) What verbs have seen the largest increase during the past 500 years? The past 100 years?
- 3) Which general classes of *se* have experienced the largest increase during the past 300 years (e.g. true reflexives: *Juan se vio en el espejo*; inchoatives: *Juan se fue a Madrid*; decausatives: *se quemó el pan*; or passives: *se construyó el puente*)?
- 4) Which *se* verbs are most likely to permit the suppression of *se* with causative verbs (e.g. *la hicieron sentarse* > *la hicieron sentar*), and why? ■

Decorte, Steven (Leiden)

Linguistic annotation in the Integrated Language Database of 8th-21st-Century Dutch (ILD) (A4.1)

The ILD project aims at creating a database covering the oldest up to the most recent Dutch language, in order to function as an instrument for research into the Dutch language and culture throughout the centuries (cf. Kruyt 2000). It will consist of three components covering all periods: a dictionary component, a corpus component and a lexicon component (electronic word lists with machine-readable linguistic information). Links within and between these components – e.g. between the dictionary quotations and the original sources or between the dictionaries themselves – make possible a wide range of research possibilities. In order to establish these links – and, of course, to enable linguistically sound retrieval – all tokens of the corpus texts will be annotated for Part of Speech (following the EAGLES/PAROLE standard; cf. Dutilh & Kruyt 2002) and headword. Only one tagset and tag method will be applied for the entire diachronic corpus. My presentation will focus on the tagging of phenomena of language change throughout the corpus. Therefor the EAGLES-term ‘transcategorization’ will be considered first, after which intended solutions will be discussed to combine ‘lexical’ tagging (lexicon-based) and ‘functional’ tagging (based on the grammatical function of the token). Structural changes on these fields – so called lexicalization and grammaticalization – necessitate the use of dictionaries in the tag process, introduction or omission of attributes and parts of speech in the tagset, extra combi-native codes and rules of priority. ■

Dufresne, Monique / Dupuis, Fernande / Tremblay, M. (Montréal)

La cartographie des préverbes en français médiéval (B5.2)

Le français médiéval est caractérisé par un processus de préfixation verbale extrêmement productif. Nos premières recherches sur des corpus de textes de cette période confirment que ces préfixes ne forment pas une classe homogène. Ainsi, si certains de ces préfixes comme *par-* et *re-* affectent surtout la valeur aspectuelle du verbe, d'autres tels *entre-*, *a-*, *en-* ont aussi un effet sur la valence du verbe. Par ailleurs, ces préfixes ont des statuts morphosyntaxiques distincts : certains dont *a-*, *con-* sont inséparables, alors que d'autres manifestent plus de liberté en ce qu'ils peuvent être séparés de la racine verbale par un clitique, un modal ou un auxiliaire. Finalement, l'hétérogénéité de ces préfixes se reflète dans leur productivité ; par exemple contrairement aux préfixes *a-* ou *entre-* qui demeurent productifs jusqu'au 16^e siècle *con-* ou *par-* se lexicalisent dès le 13^e siècle. Notre communication poursuit deux buts. Dans un premier temps, nous présenterons une description exhaustive des propriétés sémantiques et morphosyntaxiques des préfixes en français médiéval, description qui nous permettra d'inférer le système de la préfixation du très ancien français. Dans un deuxième temps, nous formulerons des hypothèses sur le rôle des processus de lexicalisation dans l'érosion du système préfixal médiéval. Celles-ci prennent en compte des contraintes morphosyntaxiques, dont la grammaticalisation de l'aspect et la restructuration du syntagme verbal. Finalement, nous verrons que la lexicalisation des préfixes tire son origine de celle des prépositions incorporées en bas latin. ■

Eggert, Elmar (Bochum)

Analyse eines Korpus zu französischen Wohnernamen (*gentilés*). Die Anwendbarkeit sprachlicher Regelmäßigkeiten in der automatischen Sprachverarbeitung (A3.2)

Der Vortrag zeigt anhand des heterogenen Gebiets der französischen Toponymenderivate (*Lyon* > *Lyonnais*, *Charleville* > *Carolo-politain*, *Pau* > *Palois*) auf, wie stark regelgeleitet die Ableitung in diesem Bereich ist und wie diese Regelmäßigkeit für die automatische inhaltliche Erfassung von Texten genutzt werden kann.

Da angesichts der Vielzahl von französischen Ortsnamen, die prinzipiell alle einen eigenen abgeleiteten Wohnernamen haben, nur die (zahlenmäßig wenigen) wichtigsten den meisten Franzosen vertraut sind, bilden die Sprecher zur Bezeichnung der Bewohner kleiner Gemeinden schnell eigene Ableitungen, die auf der Form des Ortsnamen basieren und regelmäßig sind – meist unbewusst. Diese Neubildungen folgen bestimmten Regelmäßigkeiten, die bereits die Ableitung des größten Teils der amtlichen Wohnernamen umgrenzen. Während allein schon die Wohnernamen aller Gemeinden in Frankreich nicht in einer Liste zur Verfügung stehen, sind die Neubildungen *per definitionem* nicht gelistet.

Der Bereich der Wohnernamen ist bisher vor allem aus einer onomastischen Perspektive untersucht worden, wobei sich die Linguisten auf eine Sammlung der bekanntesten Wohnernamen stützten. Die Erstellung eines größeren Korpus bzw. einer Datenbank von 10 000 Paaren „Ortsname – Wohnername“ eröffnet eine andere Perspektive. Gerade auch die automatische Textverarbeitung verlangt nach einer nicht einzelfallbezogenen Sprachbeschreibung, sondern nach Prinzipien, die die beobachteten Einheiten erklären und somit auch greifbar machen.

Zunächst soll das Gebiet der Wohnernamen vorgestellt und die Regelmäßigkeit dieses Bereichs verdeutlicht werden, nach der auch neue Ableitungen gebildet werden. Dafür werden die Regelmäßigkeiten aufgezeigt, die die Untersuchung des umfangreichen Korpus von ca. 10 000 Paaren „Ortsname – Wohnername“ ergeben haben. Auch Veränderungen der Form der offiziellen Wohnernamen in den letzten hundert Jahren sollen in einem begrenzten Rahmen nachgezeichnet werden und können teilweise durch die Bildungsprinzipien erklärt werden.

Anschließend soll aufgezeigt werden, mit welcher Vorgehensweise die Regelmäßigkeiten in der Untersuchung von Texten Anwendung finden.

Die Anwendung beschränkt sich auf ein eng umrissenes Gebiet, das der Ableitung von Ortsnamen. Auch sind solche Wohnernamen meist auf eine regionale bzw. lokale Verwendung beschränkt. Aber prinzipiell scheinen Regelmäßigkeiten der Sprache im gesamten Bereich der Ableitung bzw. der Wortbildung nutzbar zu sein für die Vernetzung von sprachlichen Einheiten, d.h. in der Zuordnung von zunächst unbekanntem Einheiten zu Elementen eines Textes, mit denen sie in einem bestimmten Verhältnis stehen. ■

Enrique-Arias, Andrés (Palma / Illes Balears)

Evolución histórica de la posición de los marcadores de objeto en español desde una perspectiva tipológico funcional: un estudio de corpus (B12.2)

Los diferentes patrones de posición de los pronombres clíticos del español a lo largo de su historia han sido documentados profusamente, no obstante, aún no está claro el por qué en su evolución de pronombres átonos a afijos verbales estos elementos acabaron por fijarse en su posición actual. Este estudio hace uso de diversos corpus textuales digitales para probar la relevancia de varias generalizaciones formuladas desde una perspectiva tipológico funcional para explicar el origen y la evolución histórica de los marcadores morfológicos verbales y su posición con respecto a los lexemas a los que acompañan: la hipótesis de la sintaxis histórica (Givón 1979) según la cual los marcadores morfológicos conservan la posición de las palabras independientes de las que derivan; la hipótesis de la serialización natural (Hawkins y Gilligan 1988), que considera que los afijos flexivos son núcleos de palabra y tienden a ordenarse como los núcleos en la sintaxis dando lugar a más sufijos en las lenguas de núcleo final y prefijos en las lenguas de núcleo inicial; y la propuesta de Enrique-Arias (2002) según la cual los marcadores de concordancia exhiben universalmente mayores porcentajes de prefijación para evitar secuencias demasiado largas de morfemas al final de las raíces verbales. Además se considera el factor psicolingüístico conocido como “priming effect” (Scherre y Naro 1991), que favorece la existencia de estructuras paralelas y según el cual la posición de los clíticos en contextos variables se vería influida por la posición de otros clíticos que aparecen cercanos en el discurso. ■

Eufe, Rembert (München)

Sprachwandel in venezianischen Kolonialverwaltungsdokumenten um 1500 (B4.2)

Ausgangspunkt meines Beitrages zum Venezianischen ist letztlich die Frage, ob das Venezianische tatsächlich die Verwaltungs- oder Staatssprache der Republik Venedig war, wie man es bisweilen hört. Das Venezianische wurde nie offiziell zu einer solchen erklärt, ein Blick auf die venezianische Schriftlichkeit zeigt jedoch, daß die lokale Sprachform sehr wohl für die Dokumente der staatlichen Verwaltung verwendet wurde. Am ausgiebigsten war der Gebrauch der venezianischen Skripta im 15. Jahrhundert, als einerseits das Lateinische den romanischen Sprachformen in einem nennenswerten Ausmaß gewichen war und sich andererseits das Toskanisch-Italienische noch nicht in dem Maße durchgesetzt hatte wie zu späteren Zeiten.

Gerade bei einer Skripta ist nicht mit sprachlicher Einheitlichkeit zu rechnen, sondern mit einem Gemisch verschiedener Formen und Schreibweisen. Um die Instabilität der venezianischen Skripta zu zeigen, in der die Merkmale des Toskanisch-Italienischen mit der Zeit immer zahlreicher wurden, bietet es sich an, Verwaltungstexte des 15. und 16. Jahrhunderts einander gegenüberzustellen. Zu Venedig selbst liegen schon Studien auf diesem Gebiet vor; hingegen wurden die Quellen aus den venezianischen Besitzungen in Dalmatien, Albanien oder Grie-

chenland bisher kaum aus dieser Perspektive betrachtet. Auch sie verdienen aber im Rahmen der eingangs gestellten Frage nach dem Status des Venezianischen in der Republik Venedig eine Untersuchung, denn es wäre ja denkbar, daß sich das Toskanisch-Italienische weniger oder gar nicht in den Kolonien ausgebreitet hat und sich dort – möglicherweise von offizieller Seite toleriert oder gefördert – das Venezianische hielt (Stichwort „Veneziano coloniale“). Daher habe ich aus dem Dokumentenbestand „Duca di Candia“ des Staatsarchivs von Venedig drei verschiedene Texttypen ausgewählt, nämlich *lettere missive e responsive* (Sendbriefe der örtlichen Kanzlei), *bandi* (Proklamationen) und *memoriali* (Gerichtsprotokolle). Dieses bei der Räumung Kretas vor der Eroberung durch die Osmanen nach Venedig transportierte Material ist nur mit große Lücken erhalten; als relativ vollständig erwiesen sich die Jahrgänge 1472 und 1567, die ich transkribierte und anschließend in elektronisch verfügbarer Form mithilfe der Mittel moderner Datenverarbeitung verglich. ■

Finck Ziller, Diomara (Eggenstein)

Funcionalidade e integração: Expansão e redução de cláusulas finais na fala do sul do Brasil (B16.1)

O presente trabalho tem por objetivo um estudo sintático-semântico-discursivo da expansão e redução de cláusulas finais do português na fala de Florianópolis, sul do Brasil. Para a realização deste trabalho, obtivemos dados em português a partir do *corpus* da região urbana do município de Florianópolis, Santa Catarina, pertencentes ao banco de dados do Projeto VARSUL (Variação Lingüística na Região Sul do Brasil).

O quadro teórico utilizado neste trabalho baseia-se no Funcionalismo. O Funcionalismo Lingüístico amplia os níveis de abrangência da gramática, permitindo incluir a combinação de cláusulas em um processo de gramaticalização, através do qual construções relativamente livres no discurso, motivadas pela situação de interação comunicativa, evoluem para construções relativamente fixas na gramática. Podemos incluir a combinação de cláusulas na gramaticalização, assumindo que esse processo abrange motivações para o desenvolvimento de estruturas gramaticais em geral.

Nesse sentido, apresentamos a codificação das cláusulas – reduzidas de infinitivo e desenvolvidas – em contextos adverbiais finais, em português, descrevendo, neste artigo, três fatores lingüísticos que caracterizam o uso dessas respectivas cláusulas. Averiguamos, ainda, graus de integração e possíveis estágios de gramaticalização.

Os resultados obtidos por meio da análise quantitativa (Varbrul) e qualitativa apontam, para as cláusulas finais investigadas, usos praticamente categóricos quanto à sua forma de codificação. A possibilidade de uma análise variacionista das cláusulas finais foi eliminada devido ao número reduzido de dados de uma das variantes. Além disso, os resultados indicam possibilidades de integração e de estágios diferentes de gramaticalização. ■

Fischer, Susann (Stuttgart)

The Case of Old Romance clitic placement revised: Data from Old Catalan (B12.1)

0. In this paper I present a comprehensive syntax of the personal pronoun clitic system of Old Catalan that not only allows to account for those data that do not find any explanation within the prominent analyses of Old Romance, but also accounts for the change in clitic placement from Old to Modern Catalan. The Catalan data I present is from a corpus of more than 7000 sentences that I have collected.

1. In order to account for the fact that Old Romance pronominal clitics are not allowed in sentence initial positions, and never appear postverbally in embedded sentences, the two prominent

views see clitic placement either due to the Verb Second phrase structure of the medieval Romance languages in cooperation with a phonological filter (Benincà 1995); or invoke a last resort mechanism (Lema & Rivero 1990, Cardinaletti & Roberts 1991). These analyses repeat the view of traditional Romance philologists (e.g. Meyer-Lübke 1897) who see the crucial factor which determined the distribution of the Old Romance clitics in their status as phonologically enclitic elements. Whereas the analyses presented in order to account for the change in clitic placement from Old to Modern Romance, explain this change with a change in the categorial status of the clitic itself (Barbosa 1993, Fontana 1993).

2. Old Catalan pronominal clitic elements contradict these views. With respect to the change in the categorial status from XP to X^o: Old Catalan clitics are never separated from the verb, contrary to what has been claimed for all Old Romance clitics, and they never appear left of the negation which strongly suggests that they need to be analysed as heads from the very beginning onwards. Nevertheless, they appear pre- and postverbally in matrix, as well as in embedded sentences, irrespective of whether the verb is preceded by a constituent or not. As for phonological cliticisation: they lean to their right as well as to their left, i.e., they are enclitic as well as proclitic.

3. On the basis of the data that I have collected I will show that syntax alone linearises the clitic-verb sequence, and that phonological cliticisation applies only after the linearisation between verb and clitic has been established, highly depending on whether a vowel is in direct neighbourhood or not.

A. The syntactical part of this analysis makes use of the old assumption that affirmation and negation are generated in the same position in the phrase-structure (Chomsky (1957), see also Laka (1990) and Martins (1995) for a similar proposal). A further category between C^o and T^o is assumed, namely Σ^o that is either headed by affirmation, emphasis or negation. I propose that, in the emergence of Old Catalan, verb-movement depends on the operator realised and that according to whether the verb moves or not the clitic either appears pre- or postverbally.

B. The phonological part of the analysis proposes that the phonological features of the Catalan pronominal clitics are introduced after Spell-Out by phonological operations applying on the way to PF. This allows to explain the perceived allomorphy within the clitic system that is highly dependent on the environment: whether post- or preverbally, and whether or not in the environment of a vowel.

4. The change that occurred with respect to clitic distribution from Old to Modern Catalan is due to a change in the parameter concerning verb-movement. This change can easily be accounted for if we adopt a theory on language acquisition and markedness along the lines of Roberts & Clark (1993). Under this assumption verb-movement to Σ^o disappeared because the trigger for it was no longer categorial, i.e. over a long period (16th until 17th century) the sequence verb-clitic, clitic-verb existed next to each other indicating the same meaning. After some time during which the learner had access to both possibilities he preferred the default (unmarked) parameter, i.e. the structure with less movement. As a consequence of this, the possibility of clitics to occur postverbally in finite clauses was lost in Catalan.

References:

- Barbosa, Pilar (1993), On Clitic Placement in Old Romance and European Portuguese, *CLS* 29, Vol. 1, 33-59.
- Benincà, Paola (1995), Complement Clitics in Medieval Romance: the Tobler-Mussafia Law, Adrian Battaye and Ian Roberts (eds.), *Clause Structure and Language Change*, Oxford, 325-344
- Cardinaletti, Anna and Ian Roberts (1991), Clause Structure and X-second, To appear in: G. Horrocks, *Levels of Representation*.
- Chomsky, Noam (1957), *Syntactic Structures*, The Hague.
- Laka Mugarza, Itziar Miren (1990), *Negation in Syntax: On the Nature of Functional Categories and Projections*, MIT Working Papers in Linguistics.
- Martins, Ana Maria (1995), Clitic Placement from Old to Modern European Portuguese, *Historical Linguistics*, Henning Andersen (ed.), Amsterdam, 395-307.
- Meyer-Lübke, W. (1897), Zur Stellung der tonlosen Objektspronomina, *Zeitschrift für Romanische Philologie* 21, 313-334.

Roberts, Ian and Robin Clark (1993), A Computational Model of Language Learnability and Language Change, *Linguistic Inquiry* 24:2, 299-345. ■

Forcada, Mikel L. / Garcia-Sempere, Marinela / Garrido, Alícia /
Gilabert-Zarco, Patrícia / Iturraspe Bellver, Amaia / Montserrat, Sandra (Alacant)

Desenvolupament a base de corpus d'un analitzador morfològic de català antic (A7.2)

La comunicació descriurà un projecte en marxa que reprén el desenvolupament d'un analitzador morfològic (que inclou, per tant, un lematitzador) de català antic públicament accessible per internet. Pretenem que aquest analitzador es convertisca en una eina útil per al tractament morfològic de corpus de català antic; a banda d'això, serà el primer mòdul d'un sistema futur l'objectiu del qual és oferir automàticament una lectura en català modern de qualsevol text antic en suport informàtic. El projecte es basa en un analitzador existent però de cobertura limitada (<<http://internostrum.com/lematitzador/>>) desenvolupat per alguns dels autors prenent com a base el vocabulari i paradigmes de flexió antiga recollits en un diccionari manual (Costa, M. i Tarrés, M. (1998), *Diccionari del Català Antic*, Barcelona: Ed. 62); l'analitzador existent és extremament veloç gràcies a l'ús de tècniques d'estats finits (10.000 mots per segon sobre un PC convencional). El sistema es genera automàticament a partir de les dades lingüístiques, cosa que permet una actualització contínua i senzilla del programa. A més de reorganitzar els diccionaris i els paradigmes de flexió i de variació gràfica del lematitzador actual, el projecte es proposa usar corpus de textos catalans antics públicament disponibles (RIALC, Biblioteca Virtual Joan Lluís Vives) per a fer el sistema més robust front a variacions gràfiques i criteris divergents de transcripció i per a millorar-ne la cobertura (fracció de text analitzat) mitjançant la inclusió d'entrades segons la freqüència d'aparició observada. L'objectiu final és la construcció d'un analitzador morfològic de català antic que siga ràpid, robust, lliurement accessible per internet, i fàcilment integrable en altres aplicacions (com ara els cercadors o indexadors de biblioteques digitals). La comunicació descriurà les estratègies lingüístiques i informàtiques que usa el projecte per a aconseguir aquestos fins. ■

Gallegos Shibya, Alfonso (Guadalajara / Freiburg i. Br.)

Nominalización y registro técnico. Algunas relaciones entre morfopragmática, tradiciones discursivas y desarrollo de la lengua en español (B2.2)

Este trabajo se divide en dos secciones. En la primera parte (aspecto sincrónico) se discuten los parámetros semióticos y morfopragmáticos generales que subyacen los procesos de nominalización en contextos (lenguajes) especializados, a partir de lo cual es posible afirmar que en español existe un reparto discursivo que guía la elección entre varias posibilidades simultáneas de nominalización.

En la segunda parte (aspecto diacrónico) se analiza el surgimiento y consolidación de un patrón de nominalización (-do) vinculado especialmente al registro técnico; para ello se ha considerado un corpus de textos técnicos que van desde el castellano medieval hasta el siglo XX. A partir de este análisis se demuestra 1) que la nominalización por medio del formativo -do en el registro técnico se fomenta en la época del español moderno (a partir de segunda mitad del siglo XVIII), y 2) que los derivados en -do adquieren paulatinamente propiedades nominales en un proceso paulatino que culmina con el tratamiento de objetos de segundo orden (abstractivos) como nombres de primero orden (lenguaje objeto).. ■

Gerner, Hiltrud (Nancy)

La Base des Lexiques du Moyen Français (BLMF) (A2.2)

La Base des Lexiques du Moyen Français (BLMF), dont une première version, encore inachevée, a été présentée au Colloque d'Ottawa en octobre 2002, est maintenant accessible via Internet. Grâce au cumul de treize lexiques, qui analysent des écrits de la période 1330-1500, elle constitue la première étape d'un dictionnaire électronique du moyen français. À ce jour, sa nomenclature compte 26 000 lemmes. Progressivement, cette base sera enrichie d'une deuxième série de lexiques, puis du Lexique complémentaire – en cours de rédaction – qui traite les vocables (au moins 30 000) qui ne figurent pas encore dans la nomenclature actuelle. Selon les prévisions, cette nomenclature finale atteindra 60 000 lemmes.

La base actuelle (logiciel Stella, développé par J. Dendien et adapté par G. Souvay) permet trois types d'interrogation :

- 1) dans des listes déroulantes, l'utilisateur peut choisir un lemme, une vedette, un étymon ou une graphie pour visualiser le ou les articles correspondants ; des liens établis entre ces différents accès facilitent la navigation d'une liste à l'autre.
- 2) la fonction "Recherche en texte intégral" permet de trouver un mot ou une séquence de mots dans l'ensemble du texte sans tenir compte du balisage.
- 3) un formulaire "Recherche avancée" permet une interrogation multi-critères en combinant plusieurs champs

Par ailleurs, un prototype de lemmatiseur est à l'essai ; il consiste à rechercher, dans le fichier des graphies-occurrences, pour un mot ou une séquence de mots sélectionnés par le passage du curseur, le ou les lemmes de rattachement possibles, suivis de leur code grammatical. Cet outil, encore incomplet, intégrera prochainement le dictionnaire des formes flexionnelles ; néanmoins, dès à présent, les résultats sont intéressants et ouvrent la perspective d'une base de connaissance lexicale pour les textes anciens. ■

Guillot, Céline (Lyon) / Heiden, Serge (Lyon) / Prévost, Sophie (Paris)

Atelier / Workshop: Corpus historiques – balisage, étiquetage, exploitation

Hakulinen, Soili (Tampere)

Corpus de traductions françaises du *De consolatione philosophiae* de Boèce (B3.2)

A part le corpus de la K. U. Leuven sur deux traités de rhétorique latins et leurs traductions en français, il n'y a pas à ma connaissance de corpus informatisé réunissant des traductions en français médiéval de textes latins. M'intéressant à la langue de traduction, j'ai donc décidé d'élaborer un corpus informatisé à l'aide du programme ParaConc de Michael Barlow (Université de Rice, Houston TX). Ce corpus contient la version latine du *De consolatione philosophiae* de Boèce et deux de ses dix traductions médiévales, ainsi qu'une traduction en français moderne. Les nombreuses versions en langues vulgaires de ce texte sont la preuve de la grande diffusion de ce texte au Moyen Age ; on connaît d'autre part l'influence qu'eurent sur le français, notamment sur son lexique, toutes les traductions de textes latins dans la période du moyen français. L'objectif de ce travail est d'étudier les différents types de réalisations du COD (SN, infinitifs, propositions...) dans ces deux traductions contemporaines (fin 14e siècle) et de mettre en rapport les données obtenues avec celles de la version latine et de la version en français moderne. Je

m'inspire dans ces travaux des approches adoptées par Michèle Goyens et de Willy van Hoecke,² ainsi que de celle sur la valence verbale de Lene Schøsler.³

Plus tard, ce corpus pourra être complété de façon à ce qu'il contienne, si possible, toutes les traductions médiévales françaises de *La consolatio* de Boèce. L'espace de temps recouvert par ces traductions s'étend du début du 13^e jusqu'au 15^e siècle. ■

Hinzelin, Marc-Olivier (Konstanz)

Die Untersuchung der Stellung der Objektpronomina zum finiten Verb anhand eines Korpus zur (alt)okzitanischen Syntax (B10.2)

Im Vortrag möchte ich über den Aufbau eines Korpus zum (Alt)Okzitanischen und dessen Verwendung für die Analyse der Pronominastellung berichten.

Es existieren nicht viele Korpora zum Okzitanischen und insbesondere wenige zum Altokzitanischen, auf dem hier der Schwerpunkt liegen soll. Die letzteren bestehen fast ausschließlich aus Troubadourichtung (z.B. *Trobadors. Concordanze della lirica trobadorica in CD-ROM*, 2001, <<http://www.sismel.it>>). Für eine syntaktische Untersuchung sind diese Korpora aber denkbar ungeeignet: Zum einen handelt es sich bei den Quellen oft um Abschriften älterer (verlorengegangener) Handschriften oder es wurden mehrere Handschriften für Texterstellung benutzt – bei beiden Verfahren gehen grammatische Besonderheiten des Ursprungstextes verloren oder sind nicht mehr eindeutig zeitlich und räumlich zuzuordnen. Zum anderen handelt es sich um Lyrik, in der die Wortstellung, oft allein aus ihrer Verwendung als literarisches Stilmittel oder aus Gründen des Reimes, Versmaßes und Rhythmus viel freier als in normal verwendeter Sprache ist oder gar deren Regeln bricht.

Als Quellen sind deshalb ausschließlich nicht-literarische Prosatexte ausgewählt worden. Hauptsächlich stütze ich mich auf Urkunden (*chartes*) und andere Dokumente aus den Archiven der untersuchten Gebiete. Einen sehr guten Überblick über die vorhandenen Editionen bieten Brunel (1922) und Martel (2001). Den sehr frühen Zeitraum mit den ersten Texten in der Volkssprache überhaupt deckt Brunel (1926, 1952) mit seiner Edition von *Les plus anciennes chartes en langue provençale. Recueil des pièces originales antérieures au XIII^e siècle* (Zeitraum 1034-1200) ab. Für das Korpus benutzte Editionen zu späteren Stadien des Okzitanischen sind Aubenas (1956), Brunel (1956), Royer (1977, 1988)

Die Stellung der schwachen Pronomina zum finiten Verb weist im Altokzitanischen interessante Unterschiede zum Altfranzösischen auf. Das Altokzitanische erlaubt, ähnlich dem Europäischen Portugiesisch, im Hauptsatz eine Stellung direkt nach dem finiten Verb, wenn der Satz nicht negiert ist. Anders als im Altfranzösischen gilt dies unabhängig von V1-Konstruktionen mit dem Verb in absoluter Anfangsposition. Das Neuokzitanische hingegen verhält sich in diesem Bereich wie das Neufranzösische. Dieser – nach meiner Hypothese möglicherweise durch den Sprachkontakt induzierte – Wandel soll hier betrachtet werden. ■

² Cf. par exemple "La traduction comme source pour l'étude d'anciens états de langue", *Le Moyen Français* 44-45/2000, p. 243-264.

³ Cf. par ex. Schøsler, Lene (1999) : "La valence verbale et l'identification des membres valenciels", dans : *Autour de Jacques Monfrin. Néologie et création verbale*. Actes du Colloque international, Université Mc Gill, Montréal, 7-8-9 octobre 1996, publiés par G. Di Stefano et R. M. Bidler. Montréal : Éditions CERES 39-40-41, pp. 527-554.

Jodl, Frank (Ludwigsburg)

Der Einsatz kommerzieller Übersetzungs-Software bei der Erforschung von sprachlicher Innovation und Sprachwandel: Der Modus nach ‚wenn, falls‘ im Portugiesischen, Brasilianischen und Spanischen (B16.2)

Eines der Probleme, die sich der Korpuslinguistik vor einigen Jahren noch stellten, nämlich das der Bereitstellung von Texten in elektronischer Form, hat in der heutigen Zeit sicherlich an Brisanz verloren. In diesem Beitrag soll dennoch ein Tool vorgestellt werden, das eben die Bereitstellung der Daten erleichtert, darüber hinaus aber die Möglichkeit bietet, sowohl onomasiologischen als auch semasiologischen Fragestellungen in einer ganz bestimmten, vereinfachten Weise nachzugehen. Dabei können einerseits Erkenntnisse auf der Ebene des Vergleichs unterschiedlicher Sprachen gewonnen, andererseits aber auch Tendenzen sprachlicher Innovation und Sprachwandel innerhalb einer bestimmten historischen Sprache aufgezeigt werden. Hauptsächlich letzteres wollen wir am Beispiel des Modusgebrauchs nach „wenn, falls“ im europäischen und im brasilianischen Portugiesisch sowie im Spanischen darstellen. ■

Kabatek, Johannes (Freiburg i. Br.)

How can we measure discourse traditions in historical corpora? (B11.1)

In the last years, one of the basic subjects of German research in historical Romance linguistic has been the study of textual traditions and its relationship to other historical evolutions. The concept of “discourse traditions” has been largely accepted as an interesting field for historical qualitative studies. The aim of my paper will be an attempt to answer two important questions:

- how can Corpus Linguistics, fundamentally quantitative, deal with a qualitative concept such as “discourse traditions”?
- is there a possibility to make the intuitively convincing conception of the existence of discourse traditions formally operationable?

The answers are crucial to postulate the integration of the study of “discourse traditions” in historical syntax, studies on grammaticalization and other historical investigations. ■

Kaiser, Georg A. (Konstanz)

Bibelübersetzungen als Grundlage für empirische Sprachwandeluntersuchungen (B8.1)

Die Verwendung von Bibelübersetzungen für historisch-vergleichende Studien ist vielfach kritisiert worden, da nach Meinung vieler historischer Sprachwissenschaftler diese Übersetzungen nicht als repräsentative Texte einer Sprache bzw. eines Sprachzustandes angesehen werden können. Zum einen wird eingewendet, dass die Sprache der Bibel, wie religiöse Sprache im Allgemeinen, eigenen stilistischen Gesetzen unterliegt. Zum anderen wird kritisiert, dass Bibelübersetzungen in aller Regel nicht (oder nicht nur) auf den hebräischen bzw. griechischen Originaltexten basieren, sondern andere Übersetzungen mitberücksichtigen oder sogar ausschließlich auf anderen Übersetzungen basieren. Dadurch enthalten dieser Kritik zufolge Bibelübersetzungen nicht nur die jeder Übersetzung inhärenten Interferenzen aus dem Originaltext, sondern auch Interferenzen, die im Laufe der vielen dazwischen liegenden Übersetzungen entstanden sind.

Demgegenüber soll in meinem Beitrag gezeigt werden, dass sich Bibelübersetzungen ausgezeichnet für empirische Sprachwandeluntersuchungen eignen. Wie für Übersetzungsvergleiche im Allgemeinen gilt auch hier, dass durch den Vergleich von Übersetzungen *eines* Textes aus unterschiedlichen Epochen die diachronische Entwicklungen einer Einzelsprache beson-

ders gut herausgearbeitet werden können. Gleichzeitig eignen sich Übersetzungen eines Textes in *verschiedene* Sprachen hervorragend dazu, die Unterschiede und die unterschiedlichen Entwicklungen zwischen den einzelnen Sprachen herauszuarbeiten. Der Vorteil eines Übersetzungsvergleiches besteht außerdem darin, dass – anders als bei vielen anderen vergleichenden empirischen Untersuchungen – weitgehend ausgeschlossen werden kann, dass die zu beobachtenden Variationen zwischen den untersuchten Texten auf Unterschieden des Textinhaltes oder der Textsorte beruhen. Der besondere Vorteil der Untersuchung von *Bibelübersetzungen* besteht dabei nicht nur darin, dass sehr viele Übersetzungen zur Verfügung stehen, sondern dass sehr häufig Bibelübersetzungen zu den frühesten schriftlichen Zeugnissen einer Sprache gehören. Von großer Bedeutung – insbesondere für *syntaktische* Sprachwandeluntersuchungen – ist die Tatsache, dass es sich hierbei größtenteils um Prosatexte handelt. Dadurch kann ausgeschlossen werden, dass zu beobachtende Wortstellungsvariationen zwischen den verschiedenen Übersetzungen von Reim- oder Versmaßeinflüssen herrühren.

In meinem Beitrag soll – ausgehend von der aus dem Jahre 1170 stammenden altfranzösischen Bibelübersetzung *Quatre livre des Reis* – exemplarisch anhand eines Kapitels dieses Bibeltextes der Vorteil eines Vergleichs von Bibelübersetzungen in verschiedenen romanischen Sprachen (und Sprachstufen) illustriert und zur Diskussion gestellt werden. ■

Klein-Andreu, Flora (Stony Brook NY)

El papel del 'corpus' en una investigación histórica funcional (B14.2)

La utilidad del estudio de corpus para la investigación histórica puede no resultar muy obvia. En mi opinión dependerá mucho de la postura teórica que se adopte.

En el ejemplo que expondré –mi investigación del desarrollo de las distintas maneras de usar los clíticos *le la lo* en castellano– se ha basado en un *enfoque funcional*, y el consiguiente interés en ver si las distintas maneras de usar estos pronombres en la actualidad pueden entenderse como resultado de *sucesivas reinterpretaciones históricas* del significado de estos elementos.

El interés por entender las diferencias observadas como resultado de reinterpretación lleva naturalmente a intentar identificar distintos sistemas de uso que constituyan un *continuo funcional*, que pueda entenderse también como probable continuo histórico. Y ello supone la necesidad de *aislar distintas variedades o dialectos* con respecto al uso de estos pronombres, geográficos y/o sociales, analizando cada uno por separado.

Por otra parte, el enfoque funcional supone un análisis inductivo, basado en la observación del uso de los elementos en cuestión –*le, la, lo* en este caso– en cada zona en *distintos entornos lingüísticos y/o pragmáticos*. De ahí la importancia de basar el análisis de cada variedad en un 'corpus' relativamente extenso.

En mi investigación del uso de *le la lo*, un corpus de 150 horas de conversaciones, en el que se distinguen hablantes de distinto nivel educativo en distintas zonas geográficas, permitió, en primer lugar, *identificar distintas variedades geográficas* en cuanto al uso de estos pronombres. Y esto a su vez ha hecho posible *proponer una posible conexión histórica* entre estas variedades o dialectos, al parecer acorde también con datos de otra clase –por ejemplo, las gramáticas prescriptivas de distintas épocas. Finalmente, esta propuesta resuelve aparentes contradicciones entre las observaciones tradicionales sobre el tema, basadas en general en observaciones individuales, de procedencia geográfica mixta o desconocida. ■

Klump, Andre (Mainz)

Zur Herausbildung der konsonantischen Besonderheiten des Spanischen auf der Insel Hispaniola (B9.1)

Von einigen (wenigen) sprachtheoretischen Zeugnissen abgesehen bezieht die historische Sprachforschung bzw. Sprachgeschichtsschreibung ihre Kenntnisse hinsichtlich der einstigen Aussprachegewohnheiten bestimmter Sprachgemeinschaften insbesondere aus der den historischen Texten zugrunde liegenden Graphie, die – zumindest was den Zeitraum der Eroberung und Entdeckung der Neuen Welt anbelangt – zwar recht uneinheitlich, aber dennoch (relativ) phonetisch orientiert war. Diese auf der Basis von Originalmanuskripten gewonnen Einzelfakten können dabei in einen historischen Zusammenhang gestellt werden, bei dem nicht zuletzt vereinzelte Schreibversehen bzw. Hyperkorrektismen eindeutige Rückschlüsse auf bestimmte Formen des mittel- und kurzfristigen Sprachwandels zulassen.

Im Zentrum der zu präsentierenden Studie steht die Aufgabe, die spanische Sprache in Santo Domingo, der ersten von Europäern errichteten Stadt der Neuen Welt auf der Insel Hispaniola, hinsichtlich ihrer phonetisch-phonologischen Besonderheiten im 16. und 17. Jahrhundert zu kennzeichnen. Besonderes Interesse gilt dabei den konsonantischen Fakten des *Seseo* (bzw. des *Ceceo*), der Elision oder Aspiration des auslautenden *-s* in implosiver Stellung, des *Yeísmo*, der Vertauschung der Phoneme /r/ und /l/ in implosiver Stellung und der Beibehaltung des anlautenden Hauchlauts [h] (< lat. *f-*), welche als Einzelphänomene in den meisten hispano-amerikanischen Varietäten verankert sind. In vereinter Form verkörpern sie jedoch die lautliche Gestalt des Spanischen im karibischen Raum und gelten heute gemeinhin als Indiz für die Verwandtschaft zwischen dem andalusischen und dem karibischen Spanisch.

Als Gegenstand für diese Untersuchung dient ein Textcorpus von über 500 (veröffentlichten) nahe- und distanzsprachlichen Dokumenten, die seinerzeit nahezu ausnahmslos in der ersten von Europäern errichteten Stadt der Neuen Welt verfasst wurden. Ergänzt wird diese sprachliche Betrachtung in die – gleichermaßen für das Textcorpus relevante – Edition von bislang unveröffentlichten bzw. unzureichend veröffentlichten privaten und offiziellen Dokumenten des 16. Jahrhunderts. ■

Knutsen, Anne Moseng (Oslo) / Ploog, Katja (Besançon)

La grammaticalisation de LA en (français) abidjanais : une “marque” à tout faire ? (B13.2)

L’emploi du français est aujourd’hui généralisé à Abidjan ; parmi les jeunes générations, il est même de plus en plus fréquemment acquis comme L1. Issu du contact de langues entre l’ancienne langue coloniale et la multitude de langues ethniques autochtones, l’abidjanais a ainsi développé des spécificités structurales à tous les niveaux d’analyse. A l’exemple du “marqueur” LA, nous nous interrogerons sur l’impact de ces restructurations sur le système de la langue.

L’étude des différents emplois de LA sur les plans phonologique, syntaxique et discursif nous conduit à poser l’existence au moins deux LA différents : le premier avec une valeur locative, semblable à un cas morphologique (sémantique), le second avec une valeur déictique ou discursive. Si l’analogie décrit extension de formes existantes à d’autres constructions (déjà existantes elles aussi), la grammaticalisation consiste en la construction d’un paradigme nouveau mettant en œuvre des formes existantes (en les privant d’une partie de leur autonomie sémantique ou fonctionnelle). Nous discuterons dans quelle mesure les traits fonctionnels caractéristiques de l’emploi du LA postnominal permettent de l’envisager comme paradigme de détermination supplémentaire et comment celui-ci s’articule avec l’“autre” LA, locatif : tout comme le paradigme des articles français, LA prend son origine dans un démonstratif, déictique, qui sem-

ble laisser son empreinte dans la valeur spécifique, stable à travers les quatre emplois différents de LA.

L'identification des restructurations d'un micro-système oblige à raisonner le continuum de pratiques tel qu'on les relève dans les discours – et fatalement dans les corpus d'étude. En effet, la démarche de description se trouve inlassablement confrontée à l'hétérogénéité du système linguistique, dont l'étanchéité est toute relative. Ainsi, nous terminerons notre exposé par une remise en perspective des caractéristiques abidjanaises en la comparant à celles d'autres variétés du français parlé et des langues créoles. ■

Langford, Chad (Lille) / O'Connor, Kathleen M. (Bloomington / Lille)

Micro-change in the French noun phrase: the case of pre-posed color adjectives (B13.1)

Our goal is to examine one specific case of micro-evolution in nominal Romance syntax: the decline of pre-posed color adjectives in the history of French. Attributive color adjectives in Modern French are overwhelmingly post-posed. However, in contrast with other typically post-posed adjectives (e.g., ethnic adjectives, adjectives of relation, and adjectives of shape), their pre-position remains possible in certain (highly marked) circumstances. The pre-position of these same adjectives in Old French, furthermore, was not uncommon. These facts raise questions concerning the decrease in the pre-position of color adjectives in the relatively recent past and the degree to which the decrease can be charted statistically with an eye to establishing key loci of its chronology.

We will evaluate the rate of pre-position of three frequently-occurring color adjectives (*noir* 'black', *blanc* 'white', and *vert* 'green') over a 400-year period (1600-2000). Preliminary results for the adjective *noir* ('black') show that it is between the end of the 18th century and the beginning of the 19th that the rate of pre-position first drops considerably before decreasing throughout the 19th and 20th centuries. We will argue that (1) their decrease is a reflection of the diachronic evolution of the noun phrase and (2) pre-position remains marginally possible given the frequency of color adjectives in earlier stages of French compared to that of adjectives of relation, adjectives of shape, and ethnic adjectives, all rare.

Our data are extracted from the *FRANTEXT* corpus (CNRS) and are further supported with data from the *Base de français médiéval* (UMR 8503). ■

Lavrentiev, Alexei (Novosibirsk)

Représentation de transcriptions diplomatiques de manuscrits français médiévaux dans le système XML-TEI (Projets *Charrette* et *Queste du Graal*) (A2.1)

Dans cette communication nous aborderons la question de la représentation informatisée de la transcription fidèle des manuscrits médiévaux, et en particulier celui du traitement des abréviations, des variantes calligraphiques de lettres et des cas de segmentation particulière (agglutination de mots ou déglutination de morphèmes). Les résultats que nous présentons ont été obtenus dans le cadre du Projet *Charrette* (dirigé par Prof. K.D. Uitti, Université de Princeton) et de l'édition en ligne de la *Queste du Graal* (dirigée par Prof. C. Marchello-Nizia, IUF et ENS-LSH, Lyon).

La nécessité d'encodage des marques "non-standard" et de la ponctuation des manuscrits médiévaux repose sur deux arguments. Premièrement, l'analyse de l'usage de ces marques peut avoir une valeur scientifique dans le cadre des études de syntaxe, de rapport entre l'écrit et l'oral, etc. Il peut également servir à identifier les scribes et les traditions d'écriture. Deuxièmement, les

images numériques de manuscrits, dont l'usage pourrait être considéré comme solution à ce problème, ne permettent pas d'effectuer de recherche textuelle ou d'analyses statistiques.

Le premier projet d'édition informatisée d'un texte français médiéval, qui comprenait des transcriptions diplomatiques fidèles aux manuscrits, était le Projet *Charrette*. Le choix initial dans ce projet était d'utiliser le système SGML-TEI et de recourir aux *entités* pour tous les caractères non alphabétiques ou marques diacritiques des manuscrits. L'état actuel de la technologie permet de passer au standard XML (plus facile à mettre en œuvre) et d'utiliser des *éléments* spéciaux pour les différents types d'abréviations, de marques de ponctuation, de majuscules initiales et de variantes calligraphiques de lettres tout en conservant un certain nombre d'entités particulières.

A un stade ultérieur, plusieurs feuilles de style XSLT peuvent être utilisées pour visualiser ces transcriptions XML de façon adaptée aux intérêts des chercheurs ou pour les transformer en tableaux utilisables par des systèmes d'exploitation de bases de données. Nous proposerons dans notre communication quelques exemples de ces mises en forme et de leurs utilisations possibles. ■

López, Marisol (Santiago de Compostela)

El Corpus del Gallego Actual (CORGA). Problemas de configuración y anotación (P)

Marchello-Nizia, Christiane (Paris)

Pour une linguistique historique 'outillée' (P)

Martins, Ana Maria (Lisboa)

Passive and impersonal *se* in the history of Portuguese (P)

Massip, M. Àngels (Barcelona) / Veny, Joan (Barcelona)

El Corpus de Diatopia Diacrònica de la Llengua Catalana (CDDL) (A5.2)

En aquesta comunicació volem explicar aquest projecte de Corpus que tenim a la Universitat de Barcelona, la feina que hem dut a terme fins ara i les perspectives de futur.

L'objectiu general d'aquest projecte és de reunir un corpus automatitzat de textos representatius bàsicament de la llengua no literària des del segle XIII fins al segle XX que ens permeti d'estudiar la variació diatòpica així com eventuais incidències de la diastràtica i d'intentar de trobar pautes de comportament generals de la relació entre llengua oral i escrita, a partir d'allò que s'observi en el cas concret del català.

La variació diacrònica (que requereix de tenir en compte tant la variació diatòpica com la diastràtica i la diafàsica) ens il·lustra sobre els processos de canvi i variació de les llengües modernes.

Ens trobem en la primera fase d'aquest projecte, que és la creació del corpus, però per al dialecte Balear hem iniciat ja una segona fase, el comentari lingüístic (en els aspectes de grafia, fonètica, morfosintaxi, lèxic, reflex del l'oralitat en el text). Encara no hem abordat l'exploració del corpus per a concordances, llistes, etc. Però ho comencem a organitzar perquè, en acabar el corpus, es pugui iniciar aquesta tercera fase.

En aquesta comunicació us volem fer partícips de les decisions que hem pres i dels problemes que se'ns han plantejat, en relació a

- Criteris de transcripció dels textos d'arxiu i d'homogeneïtzació dels textos ja editats
- Selecció dels textos
- Categorització de tipus de text, segons el registre i la temàtica
- Informatització del corpus
- Formalització i selecció de les dades lingüístiques per al tractament i explotació posteriors. ■

Moser, Karolin (Freiburg i. Br.)

Cortesía y descortesía en el español de Costa Rica: hacia una perspectiva diacrónica (B15.2)

Más allá de la discusión en torno a la existencia (o ausencia) de un nuevo tuteo en el español de Costa Rica, el actual paradigma de tratamiento a nivel de 2da. persona singular (que comprende formas voseantes y ustedeadas también) plantea una serie de preguntas que requieren una investigación diacrónica (Moser 2003).

Por este motivo levantamos un corpus de textos formales e informales que abarcan los siglos XVI hasta XIX. Trataremos por consiguiente de reconstruir sistemáticamente el desarrollo lingüístico de la cortesía y descortesía conforme a la génesis gradual de una variedad costarricense. Contaremos, por lo menos en parte, con resultados completamente nuevos, ya que nuestros materiales no sólo corroboran la existencia de un tuteo en el español colonial de Costa Rica (lo cual significaría que el tuteo actual no representa una innovación a nivel del sistema). Además nuestros datos ponen en tela de juicio la hipótesis de que la alternancia que existe actualmente entre formas ustedeadas y voseantes en una y la misma situación comunicativa sea reciente (Quesada 1995, Vega 1995). ■

Peemöller, Johanne (Kiel)

Graphiegeschichte und Sprachwandel: Möglichkeiten und Grenzen der Untersuchung von Sprachwandel in den ältesten Zeugnissen der französischen Sprachgeschichte (B5.1)

Auf der Basis eines Textkorpus, das von den Straßburger Eiden bis zur Legende des Heiligen Brendan reicht, soll hinterfragt werden, inwieweit die Untersuchung der Graphieentwicklung der Feststellung sprachlichen Wandels dienen kann, aber auch, inwiefern der Untersuchung durch Probleme in der historischen Überlieferung Grenzen gesetzt sind. ■

Perissinotto, Giorgio (Santa Barbara CA)

The Historical Dimension of Spanish in the Western United States: Notes for a Lexical Corpus and Bilingual Glossary (A8.1)

This communication is intended as a progress report on a project that intends to gather lexical data from historical records such as ships' manifests, inventories of religious and civic institutions, etc. which illustrate the Spanish language in use in the last half of the XVIII century in the Western United States of America. The presentation is divided into three parts: The background of the project, the proposed methodology, and a series of examples of actual documents and proposed format for the Glossary.

This paper and project have evolved from the book: Giorgio Perissinotto, ed., *Documenting Everyday Life in Early Spanish California* (Santa Barbara Trust for Historic Preservation, 1998, 405 pp.), which began as a project of documenting the material goods that were imported to California from various parts of the world via Mexico when the region was first being settled (the first European settlement in California is in 1769, in San Diego). This book is a transcription from Spanish and translation into English of 50+ documents registering the ordering and receiving of goods from various parts of the world – though through San Blas, in today Mexican State of Nayarit.

But its completion also made it clear that an extensive Spanish-English glossary of these material goods used to found, build, renovate and maintain human settlements in California and the Southwest was much needed. Intended mainly as a reference too for archeologists, anthropologists, linguists, and those involved in reconstruction and interpretation (such as the reconstruction of colonial sites by the National Park Service), the proposed Glossary would include a careful study of the documents that contain, inventories of goods ordered and received and will tell us much about these early settlers, their life and language on the far northern frontier of New Spain,

The Spanish spoken today in the same territory, while markedly different from that of the documents, is clearly indebted to these early forms. More importantly, the Glossary will document aspects of material culture that have either disappeared – linguistically as well as materially – or mutated altogether. The presentation will be illustrated with transparencies of the documents and with sample entries. ■

Rey, Christophe (Aix-en-Provence)

Les sons dans l'*Encyclopédie Méthodique* (B11.2)

La phonétique, ainsi que l'a clairement énoncé Sylvain Auroux,⁴ est le domaine des sciences du langage le plus méconnu des linguistes. Ceci est peut-être d'ailleurs dû à la connaissance tardive des sons, que nous ne pouvons véritablement situer qu'à partir du XVIIe siècle.

C'est toutefois le XVIIIe siècle qui marquera les plus grands progrès de cette discipline, dont la mise en exergue sera brillamment exécutée sous la plume des encyclopédistes.

Le *Dictionnaire universel ou raisonné* est bien évidemment l'ouvrage central de cette diffusion, mais il serait trop réducteur de ne s'en tenir qu'à lui, en passant sous silence l'apport considérable du *Dictionnaire de Grammaire & Belles Lettres* de Nicolas Beauzée et Jean-François Marmontel, contenu dans l'*Encyclopédie Méthodique*.

Notre contribution visera ainsi à dessiner les contours flous de l'évolution diachronique de l'étude des sons, à travers la maturation, par les auteurs de l'*Encyclopédie Méthodique*, des idées énoncées dans l'ouvrage de Diderot et d'Alembert. Nous dresserons également à cette occasion un tableau le plus clair possible de l'avancée de chacune des facettes de la phonétique, tant articulatoire qu'acoustique, et de la physiologie, pour évoquer en dernier lieu l'hypothèse solidement ancrée de l'émergence progressive et lente d'une discipline à part entière: la phonologie. ■

Ross, Sarah H. (Baton Rouge)

A corpus-based approach to register variation in Early Latin (B1.1)

This study examined markers of cognitive accessibility within early Latin using Givón's (1983) framework. Referring forms ranged from full NPs to reduced varieties including pronouns, bare

⁴ "Note sur les progrès de la phonétique au XVIIIe siècle", dans *Histoire des idées linguistiques* (Tome 2).

adjectives, and zero anaphora. Variation in forms was also explored using a corpus of texts representing different registers. Overall results met Givón's predictions, while significant differences also occurred across registers.

Accessibility marker variation has been found across registers in contemporary English (Biber *et al.* 1998). Particular anaphoric forms were found to encode relative distance from the antecedent. The distinctive registers of academic prose and conversation used anaphora similarly, but they also differed in which anaphoric forms were favored. To investigate Latin anaphora in light of the English data, two texts were chosen. Cato's *De Agricultura*, an overtly didactic treatise, represented academic prose, while Plautus's comedy *Amphitruo* with its lively dialog and colloquialisms was used to exemplify conversation.

Electronic texts were tagged using markup software, with three 1000-word samples from each. All referring forms were coded to indicate linguistic form (noun, pronoun, adjective, or zero), discourse function (new or given), and for the given entities, antecedent distance in NPs and expression type (repetition, synonym, exophor, or inferable).

Results revealed an overall preference for referring forms. Both registers used nouns more than pronouns and adjectives; zero anaphora varied by register. Differences were also found by discourse function. New entities generally followed Givón's predictions, with significant register variations for some forms. In given entities, nouns and zero anaphors occurred equally, but nouns encoded significantly longer distances than all other forms. Register variation also occurred in form usage. Some preference was also found by expression type, with noun anaphors typically repeating the antecedent, while bare adjectives overwhelmingly functioned as synonyms. Overall findings showed similar trends by register, while some token frequencies varied significantly by register. ■

Scarano, Antonietta (Firenze) / Signorini, Sabrina (Firenze)

Corpus linguistics e variazione diacronica. Uno studio su *corpora* dell'italiano dagli anni '60 ad oggi (A6.1)

La nostra comunicazione intende illustrare un'attività di costituzione di *corpora* di parlato spontaneo italiano e di ricerca linguistica su di essi, al cui interno sono possibili ricerche comparative di tipo diacronico sull'italiano dagli anni '60 ad oggi. In particolare, faremo riferimento alla sezione italiana del *corpus* C-ORAL-ROM, che contiene registrazioni di parlato italiano dagli Novanta in poi, e a un sottocorpus del *Corpus* LABLITA, il *corpus* Stammerjohann, prima raccolta di parlato italiano, risalente al 1965.

Il progetto europeo C-ORAL-ROM (V Programma Quadro EU IST 2000 26228, <<http://lablita.dit.unifi.it/coralrom>>) a cui partecipano, insieme a LABLITA, l'Université de Provence, il Centro de Linguística da Universidade de Lisboa, il Laboratorio de Lingüística Informática dell'Universidad Autónoma de Madrid, è finalizzato alla costituzione di un *corpus* multimediale del parlato spontaneo delle quattro principali lingue romanze (francese, italiano, portoghese e spagnolo) di circa 1.200.000 parole (300.000 parole per ogni lingua). Il progetto, iniziato nel 2001 e che si concluderà entro il 2003 ha portato a termine le attività relative alla raccolta del suono, alla trascrizione in formato CHAT, con tagging prosodico, all'allineamento testo-suono per enunciati con il programma Winpitch Corpus. Attualmente sono in corso studi linguistici comparativi, principalmente lessicali e sintattici, sulle quattro lingue romanze.

Il *corpus* Stammerjohann (raccolto a Firenze da Harro Stammerjohann), pari a circa 48 ore di registrazione, costituisce la base di dati oggetto di una ricerca nazionale appena iniziata (Ricerca FIRB "Archivi dell'italiano orale in diacronia") che prevede il recupero del materiale acustico originario tramite la digitalizzazione e l'archiviazione elettronica, la trascrizione in formato CHAT di una parte del *corpus* acustico corrispondente all'incirca a 100.000 parole, che sarà strutturata secondo i criteri sociolinguistici adottati nella strutturazione in nodi del *corpus* C-ORAL-ROM.

La comunicazione intende illustrare in particolare i criteri che assicurano la comparabilità dei due *corpora*:

- criteri diatopici (il luogo della raccolta dei dati);
- criteri di variazione diafasica (informale: familiare, privato e pubblico; formale: in contesto naturale e trasmesso);
- formato di trascrizione e sincronizzazione testo-suono-analisi. ■

Sinner, Carsten (Berlin)

Estudio longitudinal de cambios lingüísticos inducidos por contacto lingüístico (B2.1)

En Sinner (2001) presentamos la metodología y algunos de los resultados de un estudio empírico cuantitativo y cualitativo realizado en 1999 sobre la base de un corpus oral. Se trataba de un trabajo enfocado en el empleo de elementos sintácticos, morfológicos, léxicos y pragmáticos del castellano de Cataluña y en su ubicación dentro de un supuesto núcleo mínimo de elementos integrados en las normas de uso del castellano catalán.

En esta ocasión, expondremos el esbozo de un proyecto de un estudio longitudinal que tomará como punto de partida los datos obtenidos en el mencionado estudio de 1999, así como los datos obtenidos en un análisis de dos corpus orales constituidos por textos escritos por autores catalanes y no catalanes en 1995/96 (cf. Sinner 2000). El proyecto está enfocado en la observación longitudinal de cambios semánticos o cuantitativos inducidos por el contacto con el catalán o sólo explicables con él. Pretendemos obtener información acerca de cambios lingüísticos a corto y medio plazo. Se realizarán, en intervalos previamente establecidos, análisis cualitativos y cuantitativos basados en corpus orales y escritos, constituidos siguiendo los criterios de los estudios de 1995/96 y 1999.

Se aprovecha así la existencia de un muestrario extenso de particularidades del castellano de Cataluña. Se trata de un catálogo comentado de elementos calificados como parte del núcleo mínima de la norma de uso del castellano de Cataluña o como elementos con clara tendencia a incorporarse a dicho núcleo, considerados como indicios de cambios semánticos o cuantitativos explicables por el contacto lingüístico. De esta forma también se dará continuación a la metodología desarrollada y probada en el trabajo de campo y en los análisis de los datos resultantes. Las experiencias y las observaciones metodológicas hechas durante los primeros estudios nos permitirán, además, perfeccionar la metodología.

Referencias:

- Sinner, Carsten (2000): "Frecuencia de perífrasis verbales como índice de variación lingüística". *Revista de Lingüística Teórica y Aplicada* 38, 149–159.
- Sinner, Carsten (2001): *El castellano de Cataluña. Aspectos cualitativos y cuantitativos*. Potsdam: Universität Potsdam; tesis doctoral. ■

Sitaridou, Ioanna (Hamburg)

A corpus-based diachronic study of null subjects in French (B10.1)

The paper discusses the realisation and licensing mechanisms of null subjects in Old French dialects and in Old Occitan. The issues under investigation are: (a) Old French dialects and Old Occitan, albeit null-subject languages, employ different mechanisms for licensing null subjects (synchronic claim); (b) parametric change –such as the one that the null subject parameter suffered from Old to Classical French– is caused during first language acquisition where more than one grammar of the same language (or of different languages) leads to multilingualism. How-

ever, it could also be the case that multilingualism, as a result of language contact, can also cause parametric change (diachronic claim).

In this paper, the hypothesis whereby language contact is one of the decisive triggers of grammatical change is put to the test. More concretely, we set out to explore whether or not different distributional patterns for the null subjects can be detected in different varieties of French and whether those can be claimed to influence the course of change of the null-subject parameter in French (interaction of (a) and (b)).

Based on a corpus and a database (University of Hamburg, *Multilingualism as Cause and Effect of Language Change: Historical Syntax of Romance Languages*, research project 538), at this first stage, four varieties from contact areas are investigated with regard to the distribution and the licensing conditions of null subjects: Wallon (*Médecinaire liégeois*), Picard (*Conquête*), Francien (*Sept sages de Rome*), Occitan (*Chartes*). At a second stage Middle French is investigated (*Cent Nouvelles Nouvelles*). ■

Stark, Elisabeth (München)

Explaining article grammaticalization in Old Italian (B6.1)

The emergence of article systems in all Romance languages, deriving from a language without obligatory nominal determination (Latin), has led, since the early days of Romance linguistics, to the development of various hypotheses. Recently, and in a more universalistic prospective, two features of nominal determiners or noun phrases as a whole have attracted the attention of typologists and historical linguists, i.e. SPECIFICITY and COUNTABILITY. Both are more or less connected to nominal reference and textual coherence, in that SPECIFIC and/or 'countable' referents, referents classified as optimally 'perceptable', tend to be part of the main story / argumentation line of texts ('textually foregrounded'). SPECIFICITY is directly connected to *nominal determination* and can be considered the main feature governing explicit nominal determination in early stages of article grammaticalization (cf. Blazer 1979, Givón 1981, Heinz 1982, Selig 1992, Stark 2002). COUNTABILITY can be seen as one aspect of *nominal classification* and thereby of *apprehension* (cf. Seiler 1978, 1982, 1986).

The present talk is concerned with the development of about 20 *indefinite nominal determiners and quantifiers* in 9 Old Italian texts (from the end of the 13th to the 16th century). Definite and indefinite determiners are quite different means in nominal determination and nominal apprehension and do not belong to one and the same paradigm of grammatical elements (cf. Christopherson 1939, Moravcsik 1969 etc.). A separate analysis of the distribution and textual function of *indefinite determiners* in concrete texts will show that the reason of article grammaticalization in the area of nominal indefiniteness can partly be related to SPECIFICITY in the very beginning, but has to be ultimately seen in the light of re-organization of the system of nominal classification (COUNTABILITY as a new grammatical category of Romance noun phrases) in the development from Latin to Romance (cf. Lehmann 1991, Meisterfeld 1998 and 2000). This hypothesis is sustained by evidence from a database of about 9000 examples, deriving from three important electronic corpora recently set up by Italian linguists. The talk will show continuously the growing relevance of text corpora of important texts of the remote past.

Selected references

Corpora:

OVI: *Opera del Vocabolario Italiano* <<http://www.csovi.fi.cnr.it>>

BIT: *Biblioteca Telematica Italiana* <<http://www.cibit.unipi.it>>

LIZ (*Letteratura Italiana Zanichelli* 2.0)

Blazer, Emily Deering (1979): *The Historical Development of Articles in Old French*, Diss. University of Texas at Austin, Austin/Texas.

Christopherson, Paul (1939): *The Articles: A Study of Their Theory and Use in English*, Copenhagen: Munksgaard.

- Enç, Mürvet (1991): "The Semantics of Specificity", in: *Linguistic Inquiry* 22-1, 1-25.
- Givón, Talmy (1981): "On the development of *one* as an indefinite marker", in: *Folia Linguistica Historica* II, 35-53.
- Haspelmath, Martin (1997): *Indefinite Pronouns*, Oxford: Clarendon Press.
- Hawkins, John A. (1978): *Definiteness and Indefiniteness. A Study in Reference and Grammaticality Prediction*, London: Croom Helm.
- Heim, Irene (1988): *The Semantics of Definite and Indefinite Noun Phrases*, New York: Garland Publishing, Inc.
- Heinz, Sieglinde (1982): *Determination und Re-präsentation im Altfranzösischen*, München: Fink.
- von Heusinger, Klaus (1997): *Salienz und Referenz: der Epsilonoperator in der Semantik der Nominalphrase und anaphorischer Pronomen*, Berlin: Akademie Verlag.
- Kamp, Hans / Reyle, Uwe (1993): *From Discourse to Logic. Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Dordrecht: Kluwer.
- Lehmann, Christian (1991): "The Latin Nominal Group in a Typological Perspective", in: Coleman, Robert (Hrsg.): *New Studies in Latin Linguistics. Selected Papers from the 4th International Colloquium on Latin Linguistics, Cambridge, April 1987*, Amsterdam/Philadelphia: Benjamins, 203-232.
- Lyons, Christopher (1999): *Definiteness*, Cambridge: Cambridge University Press.
- Meisterfeld, Reinhard (1998): *Numerus und Nominalaspekt. Eine Studie zur romanischen Apprehension*, Tübingen: Niemeyer.
- Meisterfeld, Reinhard (2000): „Die unbestimmte Bestimmung: Zur Entstehung des unbestimmten Artikels in den romanischen Sprachen“, in: Staib, Bruno (Hrsg.): *Linguistica romanica et indiana. Festschrift für Wolf Dietrich zum 60. Geburtstag*, Tübingen, Narr, 303-332.
- Moravcsik, Edith A. (1969): „Determination“, in: *Working Papers on Language Universals* 1, 63-98.
- Seiler, Hansjakob (1978): „Determination: A Functional Dimension for Interlanguage Comparison“, in: Seiler, Hansjakob (Hrsg.): *Language Universals. Papers from the Conference held at Gummersbach/Cologne, Germany, October 3-8, 1976*, Tübingen: Narr, 301-328.
- Seiler, Hansjakob (1982): „Das sprachliche Erfassen von Gegenständen (Apprehension)“, in: Seiler, Hansjakob / Lehmann, Christian (Hrsg.): *Apprehension. Das sprachliche Erfassen von Gegenständen. Teil I: Bereich und Ordnung der Phänomene*, Tübingen: Narr, 3-11.
- Seiler, Hansjakob (1986): *Apprehension. Language, Object, and Order. Part III: The Universal Dimension of Apprehension*, Tübingen: Narr (= Language Universals Series, 1,3).
- Selig, Maria (1992): *Die Entwicklung der Nominaldeterminanten im Spätlatein. Romanischer Sprachwandel und lateinische Schriftlichkeit*, Tübingen: Gunter Narr Verlag.
- Stark, Elisabeth (2002): "Indefiniteness and Specificity in Old Italian Texts", in: *Journal of Semantics* 19-3, 315-332. ■

Stein, Peter (Berlin)

Die Übersetzungen von Livius' 'Ab Urbe condita' in die romanischen Sprachen, 1323 bis heute. Vorstellung des Textcorpus und der Analysemöglichkeiten (A3.1)

Das Corpus ist Grundlage für die Monographie *Untersuchungen zur Verbalsyntax der Liviusübersetzungen in die romanischen Sprachen. Ein Versuch zur Anwendung quantitativer Methoden in der historisch-vergleichenden Syntax* (Tübingen: Niemeyer 1997). Diese Arbeit geht von quantitativen Daten aus, um diese dann qualitativ zu analysieren. Sie ist diachron und innerhalb der romanischen Sprachen sprachvergleichend ausgerichtet. Gegenstand sind Aspekte der Verbalsyntax der romanischen Sprachen.

In dem Beitrag sollen das Corpus und sein Zustandekommen vorgestellt werden. Die Datenanalyse für die Ermittlung der quantitativen Daten soll kritisch nachvollzogen und zur Diskussion gestellt werden. Weitere Möglichkeiten zur Auswertung des Corpus und zur Anwendung der bisherigen Vorgehensweise sollen überprüft werden. ■

Thun, Harald (Kiel)

Corpora der gesprochenen Sprache und Texte der literarisierten Mündlichkeit als Zeugnisse des Sprachwandels (B7.1)

Texte der literarisierten Mündlichkeit in der Tradition des lateinamerikanischen *costumbrismo* empfehlen sich durch Anhäufung von Merkmalen, die als regionaltypisch gelten und damit variationsträchtig sind. Sie sind daher einerseits ergiebiger als weniger elaborierte Texte (Sprachaufnahmen), andererseits weniger zuverlässig, da sie nicht selten auf Eigenheiten eines "lenguaje imaginario" beharren. Es ist jedoch schwierig, eine scharfe Trennungslinie zwischen beiden Texttypen zu ziehen. ■

Videsott, Paul (Innsbruck)

Skriptometrische Analyse Nordostitaliens: einige Fallbeispiele (B4.1)

Unser Beitrag gliedert sich in drei Teile:

Im ersten soll die skriptometrische Methode definiert werden als Verfahren, das die Modellierung und Klassifizierung der Ergebnisse der Skriptaforchung mit den Mitteln der numerischen Taxonomie – anhand großer Datenmengen – zum Ziel hat. Des Weiteren soll der Aufbau einer skriptometrischen Matrix $N * p$ erläutert werden: N Objekte (= altnordostitalienische Texte) * p Merkmale (graphische Realisierungen von phonetischen und morphologischen Dialektcharakteristika).

Im zweiten Teil soll unsere konkrete Materialbasis bezüglich Nordostitalien vorgestellt werden: 9 Meßpunkte, 315 Merkmale, ca. 200 analysierte Texte mit insgesamt ca. 100.000 untersuchten Wörtern.

Im dritten Teil sollen – nach dem Vorbild von Goebel (1979) – einige exemplarische skriptometrische Analysen durchgeführt werden: Berechnung und zeitlicher Verlauf der relativen Frequenzen ausgewählter Charakteristika, Berechnung ihrer theoretischen Frequenzen, Berechnung und zeitlicher Verlauf der effektiven Abweichungen, Ähnlichkeitsprofil.

Literaturhinweise

Dees, Anthonij (1980): *Atlas des formes et des constructions des chartes françaises du 13^e siècle*. Tübingen: Niemeyer [Beihefte zur Zs. Rom. Phil., 178].

Goebel, Hans (1979): Verba volant, scripta manent. Quelques remarques à propos de la scripta normande, RLiR 43, 344-399. ■

Völker, Harald (Berlin)

Non-lineare Hypertextstrukturen in Textkorpora (A10.2)

„Hypertext“ wird heute weitgehend gleichgesetzt mit [haipertekst]. Und wer [haipertekst] sagt, denkt automatisch an das Internet. Gerne wird darüber indes übersehen, dass sich hinter dem [haipertekst] ein [hypertekst] verbirgt, der sich durch Relationalität und andere strukturelle Eigenschaften auszeichnet, die nicht zwangsläufig an das Medium Internet gebunden sind.

Ausgehend von einigen frühen Arbeiten zum Hypertextbegriff und aufbauend auf die Arbeit mit einem historisch-romanistischen Textkorpora sollen diese Eigenschaften in Erinnerung gerufen werden. Ziel ist dabei, die Eignung hypertextueller Strukturen zur Modellierung historischer natürlicher Sprache zu diskutieren. Besonderer Schwerpunkt wird auf die empirische Erfassung des Zusammenwirkens zwischen außersprachlichen und innersprachlichen Strukturen und Regelmäßigkeiten gelegt. ■

Wanner, Dieter (Columbus)

Data ex machina: searching for explanations in Spanish diachrony (P)

Wesch, Andreas (Köln)

Textos jurídicos y administrativos españoles del primer período de la expansión colonial (B9.2)

1. Para los documentos pertenecientes al ‘universo discursivo’ de la administración y la jurisprudencia coloniales (Canarias y América) de las últimas décadas del siglo XV y de los siglos XVI y XVII, que se toman en consideración en este estudio, existe una tipología (diplomática) establecida, dentro de la cual los tipos de documentos tienen sus denominaciones determinadas: *ordenanza*, *relación geográfica*, *instrucción*, etc. El interés (lingüístico) está centrado en el contexto socio-pragmático de los textos en cuestión, y en este sentido la diplomática, que de cierta manera siempre ha sido lingüística textual *ante litteram*, ofrece aspectos interesantes, p. ej. para determinar quién se dirige a quién en los textos. Estos se analizan, pues, como entes lingüísticos situados en un marco pragmático-social definido y cuya estructuración depende de una serie de condiciones comunicativas, a las cuales las estrategias de expresión lingüística se ajustan:

- *el grado de publicidad del acto comunicativo (privado o público) y el número de interlocutores (un grupo delimitado o el pueblo íntegro)*
- *el grado de familiaridad entre los interlocutores (entre mucha y muy poca familiaridad)*
- *el grado de distancia social entre los interlocutores y una eventual relación jerárquica*
- *el grado de solemnidad del acto comunicativo (entre muy solemne y muy cotidiano)*

Estos continuos de condición comunicativa están ligados a distintas estrategias de expresión que se caracterizan por grados entre un polo de máxima y uno de mínima inmediatez. Partiendo de los actos ilocutivos de base, parece oportuno distinguir varios grandes grupos de documentos (por lo menos tres, con subgrupos y tipos mixtos): los *documentos dispositivos* (‘mandar’) los *documentos probatorios* (‘documentar’) y los *documentos petitorios* (‘pedir’).

2. Al realizar un acto locutivo –tanto por el medio fónico como por el gráfico–, un hablante sigue no solamente las reglas fónicas / ortográficas, gramaticales y léxicas de una lengua, sino asimismo las *tradiciones discursivas*, porque *hablar* significa ‘producir textos’. El dominio de las técnicas de expresarse en una lengua (saber idiomático), y el o el dominio de las técnicas de *expresarse en textos* (saber expresivo) son las dos formas del saber lingüístico siempre presentes. Así, las tradiciones discursivas se analizarán:

– en el campo de los indicadores pragmáticos, i.e. los verbos de acto locutivo (que permiten una primera identificación de diferentes *tipos* de acto locutivo y su respectiva finalidad), las estrategias de referencia al autor y al destinatario, las denominaciones de los tipos y subtipos textuales (que suelen utilizarse en la mayoría de los documentos, demostrando explícitamente que la distinción de varios tipos de texto existía en la conciencia de los escribanos), las indicaciones meta-textuales, las indicaciones acerca de la intención y las estructuras temporales y modales.

– en el campo de las tradiciones que rigen la estructuración lingüística del texto y su relación con su propósito: la sintaxis entre agregación e integración, la progresión temática (tema y rema), la cohesión textual (referencia interna en el texto, elementos anafóricos y catafóricos), elementos de estructuración; y otra vez las estructuras temporales y modales.

Al igual que los actos locutivos elementales, los tipos de texto se realizan según tradiciones más o menos fijas. ■



Verzeichnis der Beiträger · List of contributors · Index des auteurs

Balthasar, Lukas	1, 4
Bazzanella, Carla	1
Becker, Martin	2
Beeching, Kate	3
Bellini, Daniele	4
Bert, Michel	4
Bick, Eckhard	5
Bosco, Cristina	1
Bourava, Viara	5
Brasseur, Patrice	6
Clavería Nadal, Gloria	7
Costa Giménez, Elisabet	7
Davies, Mark	8, 9
Decorte, Steven	9
Dufresne, Monique	10
Dupuis, Fernande	10
Dzian, Elodie	1
Eggert, Elmar	10
Enrique-Arias, Andrés	11
Eufe, Rembert	11
Finck Ziller, Diomara	12
Fischer, Susann	12
Forcada, Mikel L.	14
Gallegos Shibya, Alfonso	14
García Sempere, Marinela	14
Garrido, Alicia	14
Gerner, Hiltrud	15
Gilabert-Zarco, Patrícia	14
Gleißgen, Martin-Dietrich	2
Guarro, Beatriu	7
Guillot, Céline	15
Hakulinen, Soili	15
Heiden, Serge	1, 15
Hinzelin, Marc-Olivier	16
Iturraspe Bellver, Amaia	14
Jodl, Frank	17
Kabatek, Johannes	17
Kaiser, Georg A.	17
Klein-Andreu, Flora	18
Klump, Andre	19
Knutsen, Anne Moseng	19
Kunstmann, Pierre	2
Langford, Chad	20
Lavrentiev, Alexei	20
López, Marisol	21
Marchello-Nizia, Christiane	21
Martínez i Taberner, Caterina	7
Martins, Ana Maria	21
Massip, Àngels	21
Matthey, Anne-Christelle	2
Módolo, Marcelo	5

Molina Compte, Caterina	7
Mondada, Lorenza	4
Montserrat, Sandra	14
Moser, Karolin	22
O'Connor, Kathleen	20
Peemöller, Johanne	22
Perissinotto, Giorgio	22
Ploog, Katja	19
Pons, Lúdia	7
Prévost, Sophie	15
Rey, Christophe	23
Ross, Sarah H.	23
Sayós i Santigosa, Rosa	7
Scarano, Antonietta	24
Schneider, Stefan	4
Signorini, Sabrina	24
Sinner, Carsten	25
Sitaridou, Ioanna	25
Slobbe, Bianca	5
Stark, Elisabeth	26
Stein, Achim	2
Stein, Peter	27
Thun, Harald	28
Tisselli, Eugenio	7
Torruella Casañas, Joan	7
Tremblay, M.	10
Veny, Joan	21
Videsott, Paul	28
Völker, Harald	28
Wanner, Dieter	29
Wesch, Andreas	29