

CIPM - UM *CORPUS* INFORMATIZADO DO PORTUGUÊS MEDIEVAL

M. Francisca XAVIER
M. Teresa BROCARDO
M. da Graça VICENTE

FCSH - UNL

Embora o português europeu, a par do galego, seja apontado como a língua historicamente mais conservadora de entre as línguas românicas, o distanciamento que os falantes portugueses de hoje sentem relativamente à língua dos textos dos séculos XIII a XV, mesmo depois de afastadas as diferenças de natureza ortográfica, justificaria, só por si, o estudo da variação e da mudança linguísticas daquele período.

Decidimos iniciar a informatização de textos medievais portugueses com o objectivo de dispormos dos dados indispensáveis para a investigação linguística sobre a fase mais antiga do português, não esquecendo, porém, que os mesmos constituem informação útil para estudos de vária natureza.

Com efeito, a constituição de um *corpus* informatizado de textos medievais portugueses que permita aceder rapidamente e com segurança aos dados de que necessitamos, quer para a investigação que torne possível contribuir para a caracterização da gramática daquela fase histórica do português, quer para poder precisar quantitativamente a variação sincrónica que, como é sabido, se relaciona estreitamente com a mudança linguística, tem sido, ao longo deste ano, uma tarefa prioritária para o Grupo do Projecto **A Gramática do Português Medieval - Contributos para a sua caracterização**, subsidiado pela JNICT

O *corpus* informatizado de textos escritos em português medieval - CIPM - está a ser desenvolvido na Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, e teve como ponto de partida o Arquivo de Textos do Português Antigo (AOPT - Archive of Old Portuguese Texts) de Stephen Parkinson (1983).

Enquanto o AOPT foi criado para permitir estudos de aspectos de natureza fonológica, que constituem o campo de interesse de Stephen Parkinson, o CIPM está a ser concebido de maneira a possibilitar estudos linguísticos em geral.

Reconhecendo-se ser também da maior importância para o estudo da gramática daquele período histórico formular hipóteses que possam assentar em textos médio-latinos anteriores ao séc. XIII, está a ser, paralelamente, informatizado e analisado, na linha de Roger Wright (1982, 1991), por António Emiliano, um conjunto extenso de textos necessariamente úteis para o aprofundamento do conhecimento sobre a origem da língua portuguesa.

De facto, o interesse dos *corpora* informatizados com vista à investigação linguística diacrónica é actualmente incontestável, justificando decisões rápidas sobre a sua realização. No entanto, a sua concretização exige um trabalho prévio, por um lado, quanto à selecção dos textos e à adaptação dos critérios das edições, e, simultaneamente, quanto aos possíveis tratamentos automáticos a desenvolver ou a adaptar para os fins em vista. Efectivamente, perante a variedade e complexidade dos dados encontrados num *corpus* escrito, é ainda imprescindível dispor de hipóteses teóricas consistentes, tanto para a caracterização de gramática/gramáticas, histórica/s ou não, como para o estudo do mecanismo da mudança linguística.

Um grande número de estudos diacrónicos, levados a cabo nos últimos anos no âmbito da

gramática generativa, comprovam o interesse da investigação linguística histórica a par da investigação sobre as diferentes línguas actualmente existentes.

Assim, encontramos frequentemente na literatura linguística recente dados de fases antigas das línguas que comprovam hipóteses avançadas para dados linguísticos actuais.

Confrontados, porém, com duas perspectivas sobre mudança linguística que nos últimos anos se têm desenvolvido no interior da teoria generativa, uma sustentada por Lightfoot (1979, 1988, 1989, 1990, 1994) e a outra por Kroch (1982, 1989a, 1989b, 1989c, 1994a, 1994b), somos tentados a investir esforços no sentido de conseguir que o *corpus* informatizado possa ser utilizado com suporte lógico específico para a verificação das propostas teóricas por eles avançadas. Só assim será possível proceder com maior rapidez e segurança a estudos semelhantes aos que se têm vindo a realizar relativamente ao inglês, ao francês, ao *yiddish* e ao castelhano, por exemplo - para o inglês, por Anthony Kroch; para o inglês e o francês, por Ian Roberts (1985, 1993, 1994); para o francês, por Marianne Adams (1987a, 1987b) e Barbara Vance (1992); para o *yiddish*, por Beatrice Santorini (1989, 1992, 1993); para o castelhano, por Joseph Fontana (1993), etc.

A informatização de um *corpus* extenso e representativo estilística e cronologicamente, bem como o suporte lógico a desenvolver, deverão facilitar a descrição dos dados existentes nos textos antigos, permitindo, a médio prazo, cálculos quantitativos e estudos estatísticos que forneçam informação objectiva sobre o português medieval, tendências de mudança e padrões de variação.

A perspectiva de Lightfoot sobre mudança linguística tem origem nos trabalhos de Halle (1962) e Anderson (1973). Segundo estes autores, a mudança é um processo de reanálise gramatical brusca, resultante da aquisição da primeira língua. As crianças desenvolvem

uma gramática que é diferente em alguns aspectos da gramática da geração dos seus pais. E as diferenças são explicadas por uma mudança no modo como são fixados os parâmetros da Gramática Universal. Por exemplo, no inglês antigo, o Parâmetro do Sujeito Nulo tinha sido fixado positivamente, sendo permitida a inversão livre do sujeito, e o Parâmetro de Direccionalidade de Regência foi fixado para a regência à esquerda, originando a ordem (S)OV. Ou em termos do Programa Minimalista de Chomsky (1993), podemos dizer que a alteração dos traços fortes *versus* fracos dos núcleos funcionais seria reponsável pela sintaxe visível daquele período.

Como sabemos, a gramática do inglês antigo sofreu mudanças importantes ao longo da história, que são consideradas catastróficas por Lightfoot, dando origem a uma língua de sujeito obrigatório, com a ordem rígida SVO.

No entanto, as alternativas linguísticas encontradas nos textos ao longo de séculos estariam relacionadas com o facto de gerações de falantes com gramáticas diferentes coexistirem durante o período de reanálise, o que explicaria, por exemplo, a ocorrência das ordens de constituintes alternativas - SOV e SVO - no inglês antigo tardio e ainda no inglês médio.

Kroch, por seu lado, seguindo Weinreich, Labov e Herzog (1968), defende uma perspectiva variacionista sobre a mudança linguística.

Segundo Kroch, a mudança é gradual e pode resultar da competição sincrónica de alternativas linguísticas, não apenas dialectais em sentido amplo, mas também coexistentes na gramática dos falantes. Segundo esta perspectiva, os falantes de uma geração que adquiriu a nova gramática com a ordem SVO pode, simultaneamente, realizar estruturas com a ordem SOV, que corresponde à gramática da geração anterior, o que revela que é

possível dispor de dois sistemas de uma mesma língua.

Assim, de acordo com a proposta de Kroch, o cálculo da alteração das frequências relativas dos dois sistemas em competição, feito com base nos dados expressos durante o período de mudança, revela que a gramática anterior vai dando lugar, a um ritmo constante, à nova gramática. No momento em que não haja evidência de dados linguísticos da outra gramática, as gerações seguintes adquirem apenas o novo sistema

E, como dissemos no início, propomo-nos detectar os aspectos em mudança no português medieval com base nos textos que vamos seleccionando e informatizando

A constituição de um *corpus* de textos informatizado para estudos de uma fase passada da língua coloca, à partida, a necessidade de todo um trabalho prévio de estabelecimento de princípios e critérios de selecção e adaptação. Este trabalho foi realizado pelos membros da equipa, que, de acordo com as suas áreas de especialidade, contribuíram para a definição dos referidos critérios, sempre acompanhando a evolução do processo de informatização

Não cabendo no âmbito da nossa investigação a realização de novas edições, e uma vez que, nos últimos anos, têm vindo a ser realizados trabalhos nesta área que unanimemente se considera responderem às exigências de rigor que hoje se colocam à actividade editorial, começámos por integrar no nosso *corpus*, informatizando-os, textos já editados. A escolha dos textos teve, pois, em consideração, como se deduz do que acima foi dito, as características das respectivas edições e, em particular, a selecção de edições realizadas com o principal objectivo de poderem constituir objectos de estudos linguísticos.

Os primeiros textos a integrarem o nosso *corpus* foram, naturalmente, a *Notícia de Torto*¹

e o *Testamento de D. Afonso II* (segundo os dois manuscritos conhecidos)², que motivaram já sessões de estudo e debates no seio da equipa e com consultores do projecto.

Integram actualmente o CIPM os 168 textos editados por Maia (1986: 41-295) (entre os quais 61 textos do séc. XIII da Galiza e do Noroeste de Portugal), bem como os 34 documentos portugueses da Chancelaria de D. Afonso III editados por Duarte (1986) (edição paradiplomática)³. A constituição do *corpus* iniciou-se, portanto, com textos não literários - documentos notariais de carácter particular, no primeiro caso, documentos reais, no outro. Trata-se, em ambos os casos, como é sabido, de documentos linguísticos de extrema importância para estudos históricos do português⁴.

Uma vez que as edições utilizadas ofereciam as maiores garantias de rigor, tomou-se como princípio para a sua informatização a conservação, sempre que possível, dos critérios de transcrição adoptados pelos editores. Só em casos pontuais, que são claramente descritos, procedemos a simplificações e adaptações (como, por exemplo, na redução dos vários tipos de <s>, na utilização de caracteres especiais, símbolos, diacríticos), que foram ditadas pelas especificidades inerentes ao suporte utilizado com vista a uma utilização automática do *corpus*. Todas as adaptações feitas durante o processo de informatização foram, naturalmente, introduzidas de acordo com os mesmos critérios em cada um dos casos.

Embora as normas de transcrição utilizadas pelos diferentes editores não sejam uniformes, pode considerar-se que os critérios seguidos, sempre pautados por um princípio de conservação fiel das formas gráficas dos manuscritos⁵, se aproximam bastante no essencial. Procedemos apenas a algumas uniformizações na versão informatizada que permitiram eliminar as discrepâncias mais sensíveis entre as edições. É o caso dos erros e

lacunas resolvidos pelos copistas, que são assinalados por Duarte (1986), mas cuja indicação eliminámos na versão informatizada de acordo com o critério utilizado por Maia (1986)⁶. É o caso ainda dos critérios utilizados na divisão de palavras, diferentes nas duas edições referidas: conservando Duarte (1986), na edição paradiplomática, as fronteiras de palavra tal como ocorrem nos manuscritos, o que não acontece em Maia (1986), optámos por realizar para a edição informatizada dos textos da Chancelaria Real uma versão "operacional" em que a divisão de palavras corresponde à da edição interpretativa de Duarte (1986), que, neste aspecto, se aproxima mais de Maia (1986).

O CIPM integrará ainda, para o séc. XIII, textos não-literários de outros tipos, como, por exemplo, a versão portuguesa do *Foro Real* de Afonso X⁷ (e também os *Tempos dos Pretos*, um pequeno texto de Jacobo de Junta, cuja versão portuguesa se enquadra igualmente neste âmbito cronológico)⁸.

No que diz respeito aos sécs. XIV e XV, com uma documentação remanescente bem mais vasta e diversificada, irão sendo integrados no *corpus* textos não-literários e literários de vários tipos. Estão já informatizados e integram o *corpus* dois extensos textos do séc. XV, que representam tipos com diferentes características: o *Livro das Três Virtudes*, também conhecido por *Espelho de Cristina* (tradução de *Le Livre des Trois Vertus* ou *Le Trésor de la Cité des Dames* de Christine de Pizan, edição de Crispim a publicar) e a *Crónica do Conde D. Pedro de Meneses* de Gomes Eanes de Zurara (edição de Brocardo (1994))⁹.

Uma vez que encontramos já nesta época numerosos testemunhos de textos (em particular literários) originalmente produzidos em épocas anteriores, estabelecemos como um dos critérios a ter em consideração na selecção dos textos a informatizar a maior proximidade entre a época da sua redacção primitiva e a datação dos testemunhos conservados. Não ficarão, porém, excluídos do CIPM textos apenas conservados em testemunhos tardios,

que virão também a ser incluídos no *corpus* - é o caso, por exemplo, dos textos editados sob o título "Vidas de Santos de um manuscrito alcobacense"¹⁰, cuja informatização, a cargo de Carlos Rocha, está a ser concluída. O estudo linguístico destes textos coloca, como é sabido, questões complexas, que apenas poderão ser correctamente equacionadas numa fase mais avançada do projecto, posteriormente ao tratamento linguístico dos restantes textos.

Neste momento, o CIPM tem prontos para análise os seguintes textos:

	Nº de palavras
<u>Século XIII</u>	
História do Galego-Português (Maia (1986))	28 804
Chancelaria de D. Afonso III (Duarte (1986))	16 866
Notícia de Torto (Cintra (1990))	779
Testamento de D. Afonso II (Costa (1979))	
Manuscrito L	1 432
Manuscrito T	1 437
Arquivo de Textos do Português Antigo (Parkinson) ¹¹	19 504
Total:	68 822
<u>Século XIV</u>	
História do Galego-Português (Maia (1986))	Total: 33 242
<u>Século XV</u>	
História do Galego-Português (Maia (1986))	30 162
O Livro das Três Virtudes (Crispim)	56 272
Crónica do Conde D. Pedro de Meneses	136 661
Total:	223 095

Algumas experiências de trabalho automatizado com *corpora* de textos linguisticamente analisados (anotação das partes do discurso e indicação da estrutura sintáctica) encorajam a prosseguir nessa via. É hoje indiscutível que a criação desses *corpora* linguisticamente úteis pode reduzir substancialmente o esforço inicial dos linguistas que sobre eles se debruçam.

Com efeito, por um lado, o trabalho fica feito e os resultados disponíveis para todos quantos se interessarem pelo estudo linguístico, estilístico ou literário dos textos, tornando-se assim desnecessário que cada investigador ou equipa de investigação faça de novo aquilo que outros já fizeram antes, cabendo-lhe apenas completar a análise em função dos seus interesses específicos.

Por outro lado, o desenvolvimento de suporte lógico adequado permite extrair automaticamente vários tipos de informação, reduzindo significativamente o tempo de preparação do *corpus* e de recolha de dados. A automatização torna-se, obviamente, mais necessária e premente à medida que a dimensão do *corpus* aumenta.

Segundo informação de Anthony Kroch¹², dois grupos de investigação da Universidade da Pensilvânia estão actualmente a mover-se nesta linha.

O "Penn Treebank" trabalha há mais de quatro anos sobre textos completos de inglês moderno - cerca de 4 milhões de palavras -, sendo o seu objectivo a explicitação das árvores sintácticas das frases do *corpus*.

O "Penn Parsed Corpus of Middle English" (PPCME), inspirado no projecto anterior, partiu de um *corpus* já informatizado - os fragmentos de textos de inglês médio em prosa e em verso do "Helsinki Corpus of English Texts", cuja extensão é de 608 570 palavras, e

que não contém quaisquer anotações linguísticas.

O "Penn Parsed Corpus" é coordenado por Anthony Kroch, com quem mantemos contactos desde há algum tempo, no sentido de uma cooperação que nos permitirá avançar no nosso projecto com a segurança da experiência efectiva nesta área daquele grupo de investigação. Pouparemos, assim, alguns erros e esforços que, de outro modo, seriam certamente inevitáveis. Estamos, naturalmente, conscientes de que o trabalho sobre o português coloca problemas específicos, para os quais haverá que encontrar soluções próprias.

O procedimento ideal para a anotação destes textos consistiria, segundo Kroch, em três etapas, que são, aliás, as seguidas no "Penn Treebank": (i) etiquetagem automática das partes do discurso; (ii) segmentação automatizada dos sintagmas com base na informação resultante de (i); e (iii) correcção manual mínima.

Todavia, a não padronização da pontuação e da ortografia e a maior variação na ordem das palavras impedem a adopção deste procedimento para os textos medievais.

Assim, a metodologia adoptada no PPCME consiste aproximadamente nos seguintes passos.

1. Segmentação manual do *corpus* num sistema hierárquico limitado: (i) isolamento e referenciação de orações matrizes e parentetização das encaixadas, (ii) etiquetagem para identificação do tipo de cada uma delas (adverbiais, relativas, completivas, etc.)

2. Indexação automática das orações.

3. Parentetização interna automática - 1ª fase: introdução de parênteses de abertura, parentetização e etiquetagem categorial de palavras simples facilmente

identificáveis.

4. Correção manual e introdução, também manual, dos parênteses de fechamento dos sintagmas

5 Parentetização interna automática - 2ª fase. Etiquetagem dos SN de acordo com a função

6 Introdução manual de vestígios de movimentos sintácticos e coindexação com os antecedentes. Correção de erros eventuais.

Com este procedimento, a percentagem de erros - 2-3% - é semelhante à obtida no "Penn Treebank", e estima-se que a economia de tempo seja da ordem de 50%.

Tendo em conta a fase inicial do nosso projecto, em que a primeira tarefa é a criação do próprio *corpus*, o trabalho do PPCME é para nós um programa, uma via aberta, para a qual serão necessários, naturalmente, tempo e meios de que não dispomos por enquanto.

¹ Cintra (1990: 37-41). Informatização de António Emiliano.

² Costa (1979: 307-321). Informatização de António Emiliano.

³ Informatização de Susana Pereira e Teresa Oliveira.

⁴ Seria supérfluo incluir aqui considerações sobre esta questão, que foi tratada de forma muito aprofundada nos artigos já clássicos de Cintra (1963a e 1963b), bem como no estudo que acompanha a edição de Maia (1986).

⁵ No que respeita a Duarte (1986), referimo-nos à edição paradiplomática.

⁶ Vejam-se as respectivas normas de transcrição Duarte (1986: 50) e Maia (1986: 23).

⁷ Edição de Ferreira (1987).

⁸ Edição de Ferreira (1986)

⁹ Trata-se em ambos os casos de edições realizadas directamente em suporte informático, o que possibilitou a sua integração imediata no CIPM.

¹⁰ Edição dirigida por Castro (1982-1983; 1984-1985)

¹¹ O arquivo, informatizado por Stephen Parkinson, é composto por revisões críticas de edições já publicadas e transcrições originais de documentos inéditos (v. Parkinson (1983))

¹² Seminário sobre "Tratamento automático de textos antigos para análise sintáctica", FCSH-UNL, 16-20 de Junho de 1994.

REFERÊNCIAS:

- ADAMS, M. P. (1987a) "From Old French to the Theory of pro-drop", *Natural Language and Linguistic Theory* 5.1, pp. 1-32
- ADAMS, M. P. (1987b) *Old French, Null Subjects, and Verb Second Phenomena*. University of California at Los Angeles (Ph D. dissertation).
- ANDERSON, H. (1973) "Abductive and Deductive Change", *Language* 48, pp. 765-793.
- BROCARDO, M. T. (1994) *Crónica do Conde D. Pedro de Meneses de Gomes Ianes de Zurara. Edição e Estudo*. Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa (dissertação de doutoramento).
- CASTRO, I. (ed.) (1982-1983); (1984-1985) "Vidas de Santos de um Manuscrito Alcobacense", *Revista Lusitana*, Nova Série, 4,5, Lisboa, Centro de Estudos Geográficos, pp. 5-52, 43-71.
- CHOMSKY, N. (1993) "A Minimalist Program for Linguistic Theory" in *The View from Building 20. Essays in Linguistics in Honor of Sylvain Bromberger*, Cambridge, The MIT Press, pp. 1-52.
- CINTRA, L. F. L. (1963a) "Les Anciens Textes Portugais non Littéraires. Classement et Bibliographie", *Revue de Linguistique Romane* 27, pp. 40-58.
- CINTRA, L. F. L. (1963b) "Observations sur l'Orthographe et la Langue de Quelques Textes non Littéraires Galicien-portugais de la Seconde Moitié du XIIIe Siècle", *Revue de Linguistique Romane* 27, pp. 59-77.
- CINTRA, L. F. L. (1990) "Sobre o mais Antigo Texto Não-literário Português: a 'Notícia de Torto' (Leitura Crítica, Data, Lugar de Redacção e Comentário Linguístico)", *Boletim de Filologia* 31, pp. 22-77.
- COSTA, A. J. (1979) "Os mais Antigos Documentos Escritos em Português. Revisão de um Problema Histórico-linguístico", *Revista Portuguesa de História* 17, pp. 263-340.
- CRISPIM, M. L. *Edição Crítica de O Livro das Três Virtudes de Christine de Pizan* (a publicar)
- DUARTE, L. F. (1986) *Os Documentos em Português da Chancelaria de D. Afonso III (Edição)*, Faculdade de Letras da Universidade de Lisboa (dissertação de mestrado).
- FERREIRA, J. A. (1986) "Edição e Estudo Linguístico dos 'Tempos dos Preitos'" in J.

Roudil, *Jacobo de Junta. Summa de los Nueve Tiempos de los Pleitos. Édition et Etude d'une Variation sur un Thème*, Paris, Klincksieck.

- FERREIRA, J. A. (1987) *Afonso X. Foro Real, Edição, Estudo Linguístico e Glossário*, 2 vols., Lisboa, INIC.
- FONTANA, J. (1993) *Phrase Structure and the Syntax of Clitics in the History of Spanish*, University of Pennsylvania (Ph.D. dissertation).
- HALLE, M. (1962) "Phonology in Generative Grammar", *Word* 18, pp. 54-72.
- KROCH, A. S. (1982) "Grammatical Theory and the Quantitative Study of Syntactic Change". Paper presented at the 11th NWAV Conference, Georgetown University. Published as Kroch 1989a.
- KROCH, A. S. (1989a) "Function and Grammar in the History of English: Periphrastic 'do'" in R. Fasold, D. Schiffrin (eds.) *Language Change and Variation*, Amsterdam, Benjamins, pp. 133-172.
- KROCH, A. S. (1989b) "The Loss of the Verb-second Constraint in Middle English and Middle French". Paper presented at the 18th NWAV Conference, Université de Montréal.
- KROCH, A. S. (1989c) "Reflexes of Grammar in Patterns of Language Change", *Journal of Language Variation and Change* 1.3, pp. 199-244.
- KROCH, A. S. (1994a) "Remarks on the XV / VX Alternation in Early Middle English". Paper presented to The Third Diachronic Generative Syntax Conference, VU Amsterdam.
- KROCH, A. S. (1994b) "Morphosyntactic Variation", in K. Beals et al. (eds.) *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory (to appear)*.
- LIGHTFOOT, D. (1979) *Principles of Diachronic Syntax*, Cambridge University Press.
- LIGHTFOOT, D. (1988) "Syntactic Change" in F. Newmeyer (ed.) *Linguistics: The Cambridge Survey*, Cambridge University Press.
- LIGHTFOOT, D. (1989) "The Child's Trigger Experience: Degree-0 Learnability", *Behavioral and Brain Sciences* 12.2, pp. 321-334.
- LIGHTFOOT, D. (1990) "Old Heads and New Heads" in J. Mascaró, M. Nespór (eds.) *Grammar in Progress: GLOW Essays for Henk van Riemsdijk*, Dordrecht, Foris.
- LIGHTFOOT, D. (1994) "Shifting Triggers and Diachronic Reanalyses". Paper presented

to the Third Diachronic Generative Syntax Conference, VU Amsterdam.

- MAIA, C. A. (1986) *História do Galego-português. Estado Linguístico da Galiza e do Noroeste de Portugal do Século XIII ao Século XVI*, Coimbra, INIC.
- PARKINSON, S. (1983) "Um Arquivo Computorizado de Textos Medievais Portugueses", *Boletim de Filologia* 28, pp.241-252.
- ROBERTS, I. G. (1985a) "Agreement Parameters and the Development of English Modal Auxiliaries", *Natural Language and Linguistic Theory* 3 1, pp. 21-57.
- ROBERTS, I. G. (1985b) "Oblique Case in the History of English", *University of Southern California Working Papers in Linguistics*.
- ROBERTS, I. G. (1993) *Verbs and Diachronic Syntax*, Dordrecht, Kluwer.
- ROBERTS, I. G. (1994) "Cliticisation and Word Order Change in the History of English". Paper presented to The Third Diachronic Generative Syntax Conference, VU Amsterdam.
- SANTORINI, B. (1989) *The Generalization of the Verb-Second Constraint in the History of Yiddish*, University of Pennsylvania (Ph.D. dissertation).
- SANTORINI, B. (1992) "Variation and Change in Yiddish Subordinate Clause Word Order", *Natural Language and Linguistic Theory* 10, pp. 595-640.
- SANTORINI, B. (1993) "The Rate of Phrase Structure Change in the History of Yiddish", *Language Variation and Change* 5, pp. 257-283
- VANCE, B. (1992) "Subject Inversion in Old and Middle French" in A. Battye, I. Roberts (eds) *Language Change and Verbal Systems*, Oxford, Oxford University Press.
- WEINREICH, U.; W. Labov; M. I. Herzog (1968) "Empirical Foundations for a Theory of Language Change" in W. P. Lehmann; Y. Malkiel (eds.) *Directions for Historical Linguistics*, Austin, University of Texas Press, pp. 95-188.
- WRIGHT, R. (1982) *Late Latin and Early Romance in Spain and Carolingian France*, Liverpool, Cairns.
- WRIGHT, R. (ed.) (1991) *Latin and the Romance Languages in the Early Middle Ages*, London, New York, Routledge.