PAMAPLA 14

PAPERS FROM THE FOURTEENTH ANNUAL MEETING OF THE ATLANTIC PROVINCES LINGUISTIC ASSOCIATION

Memorial University of Newfoundland
St. John's, Newfoundland
November 9 – 10, 1990

ACALPA 14

ACTES DU QUATORZIEME COLLOQUE ANNUEL DE L'ASSOCIATION DE LINGUISTIQUE DES PROVINCES ATLANTIQUES

Université Memorial
Saint-Jean, Terre-Neuve
le 9 – 10 novembre 1990

EDITED BY / RÉDACTION

JIM BLACK

PAMAPLA 14

PAPERS FROM THE
FOURTEENTH ANNUAL MEETING
OF THE
ATLANTIC PROVINCES LINGUISTIC ASSOCIATION

Memorial University of Newfoundland
St. John's, Newfoundland
November 9 – 10, 1990

ACALPA 14

ACTES DU
QUATORZIEME COLLOQUE ANNUEL
DE
L'ASSOCIATION DE LINGUISTIQUE DES PROVINCES ATLANTIQUES

Université Memorial
Saint-Jean, Terre-Neuve
le 9 – 10 novembre 1990

EDITED BY / RÉDACTION

JIM BLACK

ACKNOWLEDGEMENTS

The *Atlantic Provinces Linguistic Association* acknowledges the generosity of Memorial University of Newfoundland in hosting the conference and in providing a grant towards the publication of this volume. The Association gives special thanks to Dr. Michael Staveley, Dean of Arts, for his enthusiastic support.

The fantastical beast displayed on the cover was created at the conference by Sarah Thomason, our guest speaker, and is used with her permission.

REMERCIMENTS

L'*Association de Linguistique des Provinces Atlantiques* veut bien exprimer sa reconnaissance à l'Université Memorial de Terre-Neuve pour son chaleureux accueil de cette réunion ainsi que pour la subvention qu'elle a bien voulu accorder permettant la publication des présents actes. L'Association tient surtout à remercier le doyen de la faculté des arts, Michael Staveley, de son soutien enthousiaste.

La bête fabuleuse à la couverture fut créée lors de la réunion par Sarah Thomason, notre conférencière invitée, et paraît avec sa permission.

TABLE OF CONTENTS

TABLE DES MATIÈRES

V. Motapanyane was unable because of illness to present her paper at the conference.

V. Motapanyane, malade, n'a pas pu présenter sa communication lors de la conférence.

OTHER PAPERS PRESENTED

AUTRES COMMUNICATIONS PRÉSENTÉES

Rose Mary BABITCH	The Atlas of Acadian French Maritime Terminology: Progress Report
Patricia BALCOM	Argument Structure in Early Child Language
Sandra CLARKE & Deborah JAMES	Gender Differences in the Use of Interruptions: a Review
Ronald COSPER	Egyptian and Hausa: a Comparative Analysis
Cynthia GROVER	Speech and Sensation
John HEWSON	Proper Noun Inflections in Micmac
Ruth KING	Back Again
Marguerite MACKENZIE	The Sociolinguistic Situation of a Non-standard Fijian Dialect
Barry MILLER	On the Computation of Conceptual Structure
Raymond MOUGEON & Édouard BÉNIAK	Issues in the Study of Linguistic Variation in Contact Settings

Harold PADDOCK	A Case for Case Studies (of Linguistic vs. Non-linguistic Conditioning)
Shana POPLACK (guest speaker / conférencière invitée)	Black English in the Diaspora: Reconstructing Diachrony from Synchronic Evidence
Amanda POUNDER	Morphology, the Lexicon and the Rise of the (Written) Standard
Marie-Lucie TARPENT	The Evolution of the Tsimshianic Stress-Rule
Sarah THOMASON (guest speaker / conférencière invitée)	Coping with Partial Information in Historical Linguistics & Using Linguistic Clues in Solving Historical Mysteries

CONTACTS ET CARACTÈRES COMMUNS DU FRANÇAIS ACADIEN ET DU CRÉOLE FRANÇAIS EN LOUISIANE

Catherine Bodin
Mount St. Mary's College
Emmitsburg, MD, Etats-Unis

RÉSUMÉ

Il est difficile de déterminer la durée du contact acadien-créole en Louisiane, mais les deux parlers se jouxtent au moins depuis la libération des esclaves afro-américains en 1863. Depuis lors, les parlers acadien louisianais (AcL) et créole louisianais (CrL) s'influencent et s'assimilent jusqu'au point où l'AcL est vu (à tort) comme étant une créolisation du français colonial du dix-septième siècle . La créolisation s'opère depuis un siècle au moins, mais les plus précoces attestations des deux dialectes font voir l'influence du français maritime (FrMar) qui leur est commune; ce registre du français régional (FrReg) est composé des restes dialectaux du normand et du pictavo-saintongeais et comporte un grand nombre de termes maritimes. Le lexique acadien des Provinces Maritimes et les atlas linguistiques du Poitou et de la Louisiane nous aident dans la reconstruction du contact de l'AcL et du CrL, dans l'effort d'éclaircir leurs évolutions, qui continuent jusqu'au présent.

SIGLES EMPLOYÉS:

AcL	Le dialecte acadien louisianais
AcM	Le dialecte acadien des Provinces Maritimes
ALF	Atlas Linguistique de France (Gilliéron)
ALO	Atlas Linguistique de l'Ouest (Massignon)
CrL	Le créole français louisianais
FrMar	Le français maritime (16^{e}–18^{e} siècles)
FrReg	Le français régional
FrL	Le français colonial en Louisiane
FrSt	Le français standard

I. SITUATION HISTORIQUE DES DEUX GROUPES

Le créole louisianais (CrL) est de base française et date du dix-huitième siècle, présumément dès l'arrivée des esclaves africains en territoire louisianais, en 1719, à la fondation de la ville de la Nouvelle Orléans. A l'origine une langue de contact, le CrL a assimilé bien des caractères du premier contact français, sans doute le français maritime (FrMar). Or, le CrL n'est pas parlé seulement par les Louisianais noirs, les descendants des esclaves des dix-huitième et dix-neuvième siècles: bien des blancs, en particulier des Acadiens, le parlent et la langue acadienne louisianaise (AcL) se créolise, comme la langue américaine de ce siècle. L'objet de cette présentation est d'examiner les contacts des deux

langues, car les dialectalismes du FrMar et d'autres dialectalismes entrent dans les deux indépendamment de leur contact mutuel.

Les deux groupes s'établissent en Louisiane à des dates différentes, en 1719 et environ 1765 pour les Acadiens. On a raison de croire aux distinctions rigides de race et de classe dans la Louisiane coloniale. En principe, ces deux populations n'ont pas eu de contact. Les noirs libres étaient à la Nouvelle Orleans, et les Acadiens non. La majorité de noirs étaient des esclaves, et les gens de situation humble ne possédaient pas d'esclaves. D'ailleurs, la majorité d'agriculteurs n'ont pas eu d'esclaves à la veille de la guerre civile en 1860, et la plupart des esclaves étaient concentrés dans les grandes propriétés où les petits agriculteurs ne figuraient pas comme voisins. Il est vrai que le paysage change beaucoup en cent ans, depuis 1765. Mais la Louisiane est toujours une société de classe. Dès le début, les Acadiens se sont infiltrés vers l'intérieur, car ils sont arrivés par des miliers entre 1758 et 1785. Ils se sont établis sur les rivières et bayous, car c'était le modèle d'urbanisation du siècle, modèle qui a survécu en quelques 'villages en ligne' jusqu'à nos jours. Mais bientôt les grands propriétaires sont venus s'accaparer de toutes les propriétés qui donnaient sur la rivière Mississippi, qui était la voie de commerce majeure.

A leur arrivée, les Acadiens ont été bien vus et mal vus. Certains Louisianais plaignaient leur pauvreté absolue tandis que d'autres critiquaient leur indolence pendant que la chaleur et la fièvre les assomaient. Après l'ère des pionniers, leur culture de paysan n'était pas comprise. Quarante ans après, C. Robin, dans son *Voyage à l'intérieur de la Louisiane* (1804: 272, 289) a décrit des Acadiens des prairies comme étant des patoisants, mais honnêtes et propres, avant de faire la remarque insolite que les enfants ne parlaient pas anglais, à la différence de leurs parents (qui avaient plus de contacts dans le monde). Il prétend que: "Les manières et coûtumes sont entièrement françaises et la langue française est restée la langue principale du pays," même en admetant un bilinguisme répandu et se plaignant que le français était une langue menacée alors. La Louisiane s'est anglicisée tout au courant du dix-neuvième siècle.

En effet, Napoleon a vendu la Louisiane aux Américains en 1803, ayant compris qu'il ne pouvait pas la défendre de l'incursion américaine, qui avait commencé même avant la vente. Dorénavant, les deux cultures s'affrontent et les planteurs français championnent la cause des francophones. Tant qu'ils auront le dessus, le français sera protégé comme langue de statut légal et langue de littérature et de culture. Il perdra ce statut en 1914, quand l'Assemblée ne publiera plus ses actes en français, et en 1920, quand la langue sera formellement défendue dans la vie publique et dans les écoles. Il redeviendra un parler légal en 1968.

Au courant des deux siècles, les Acadiens semblent se marginaliser de plus en plus. Ils sont, d'après le sociologue Glenn Conrad (1975: 4), "une culture folklorique classique," axée sur les affaires de la famille et des intimes, appartenant à la classe ouvrière ou agricole, et ayant peu de contacts extérieurs, s'assimilant peu à la culture majoritaire.

Le statut des noirs change davantage au courant des deux siècles. A l'origine esclaves ou libres, l'émancipation universelle égalise l'état des noirs mais réduit le prestige des déjà-libres dans l'ère de racisme qui s'ensuit. Le Sud est lent à se réparer et sombre dans la pauvreté et le racisme. Les lois viennent défendre le mariage entre noirs et blancs seulement après l'ère d'esclavage. Jusqu'au vingtième siecle, les lois distinguaient la race jusqu'à la quatrième génération d'intermariage, soit aux bis-grandparents. Déjà, l'intermariage a dû être commun à une certaine époque, donné le nombre de grandparents modernes qui ne veulent rien entendre parler du sujet. Les moeurs avaient changé à quelque point entretemps. On se plaît à penser que la vraie ère de racisme est arrivée avec la

défense de parler français, qui a renforcé la hiérarchisation des gens et confirmé les préjugés contre le pluralisme.

II. ORIGINES ET CONTACTS

En haut, j'ai fait référence aux contacts des deux groupes. La chronologie, a l'égal du milieu social, régit la qualité du français qui est disponible. Le francais langue nationale perd beaucoup de son fonds dialectal pendant l'époque coloniale américaine et exerce une pression standardatrice sur d'autres formes de la langue aussi. L'écart entre les registres du français diminuera progressivement dans le temps en France, alors que les parlers du Nouveau Monde figeront quelques caractères dialectaux dans leurs lexiques de base.

Le CrL, un produit des dix-huitième et dix-neuvième siècles, a dû prendre contact avec le FrMar de cette époque, et ensuite avec le français colonial en Louisiane (FrL) qui est pourtant tres proche au français moderne. Pendant l'ère amèricaine, l'importation de nouvelles cargaisons humaines a été défendue, ce qui fait que le CrL a subi cent ans de contact (au maximum) avec le FrMar, ou moins si cette traffique s'est anglicisée entretemps.

Par contre, le dialecte acadien est parti de la France du dix-septième siècle, d'environ la Loire; il a séjourné aux Provinces Maritimes pendant un siècle, entre 1640-1680 jusqu'en 1755-65. Les colons acadiens n'ont pas été un groupe homogène et le brassage de leurs parlers locaux a dû être important. Isolés dans les Maritimes, à 1000 miles du Quebec, ils ont profité de très peu de communication avec le français extérieur et leur langage a dû conserver beaucoup de dialectalismes. De brefs contacts avec les Anglais et avec les Français (dans certains cas) n'ont pas grandement influencé leur langue avant leur arrivée en Louisiane cent ans après leur émigration aux Maritimes. Le fait que l'AcL moderne donne beaucoup moins de dialectalismes que l'acadien des Maritimes (AcM) montre que le brassage qu' a subi le premier en Louisiane a été très important.

III. PROFIL LEXIQUE

Un profil lexique décrit en partie les différences entre l'AcL et le CrL; il les distance en même temps qu'il établit leurs points communs.

Un lexique abrégé est donné en apendice I. Les notes sur le CrL renvoient à trois sources, dont l'article de Hull (1968), qui a identifié comme registre le FrMar à partir des créoles de la Caribée, la Louisiane, et l'Océan Indien; aussi, Fortier a publié une collection de contes folkloriques (1894) qui donne le plus grand échantillon du lexique; finalement, Broussard (1942) a étudié cette langue en présentant des dictons, une collection unique de poésie écrite par un esclave, et des traductions de contes qu'il a lues à des Créoles qui ont approuvé son langage. Les deux dernières sources, des professeurs louisianais qui parlaient un français soigné, avaient le créole comme langue maternelle, car les enfants dans les classes privilégiées ont été élevés par les domestiques jusque dans ce siècle. Une quatrième source qui n'est pas cité ici est Morgan (1959), qui donne un créole de la paroisse de St. Martin qui a beaucoup assimilé du français, ou le FrL ou l'AcL. Morgan a fait des enregistrements sur place, dont il a publié en 1959; sur le plan chronologique, il y a une différence d'environ cinquante ans entre sa publication et les connaissances de l'enfant James Broussard, qui a dû apprendre un créole plus vieux de la bouche des adultes.

L'AcL, pour sa part, n'est pas étudié avant Fortier (1894) qui parle simplement du 'français' parlé à deux endroits sans l'identifier comme étant acadien. Son rapport sur le français parlé 'à l'intérieur' [en amont du Bayou Tèche] dans les paroisses civiles de St. Mary et de St. Martin donnent des maritimismes qui proviennent surement de l'AcL. Ces mêmes termes sont attestés dans un Atlas Linguistique de Louisiane du siècle présent. Fortier distingue le FrL de l'AcL lors de sa description de St. Martinville en 1894, mais sans donner des critères: il y va de l'accent et de l'intonation autant que de la classe de l'interlocuteur. A l'époque, la question de la correlation de classe et de dialecte ne se posait même pas; elle était considérée comme allant de soi. En finale, Fortier donne le texte d'une lettre qui lui a été addressée par des Acadiens qu'il avait connus à un bal de samedi soir et dont le langage se rapproche bien du FrSt en dépit de son orthographe.

Plus tard, un ensemble de thèses de maîtrises recherchées a l'Université de l'Etat de Louisiane et datant des années vingt jusqu'à quarante de ce siècle, prouve la correlation entre l'AcL et le français recensé par Fortier. Ces thèses sont une collection de glossaires qui présentent le français de plusieurs paroisses dans le dégré qu'il se diffère du FrSt. Ces études préscriptives offrent l'avantage de recueillir les dialectalismes et régionalismes qui, effectivement, différencient le FrL du FrSt. A un endroit seul (la paroisse d'Ibérie, (Trappey: 1940)), l'enquêteur fait-il la différence entre le FrL et 'la variété dialectale' du français local (i.e., l'AcL). La distinction à cette époque tardive est surprenante, car les enquêteurs du *Works Project Administration* de la même époque ne peuvent plus les différencier; aussi, Conwell (1975: 29), écrivant de Lafayette en 1962, dit également que le FrL et l'AcL sont assimilés.

Mes enquêtes personnelles font voir que l'AcL moderne (tel qu'il est parlé par des octogénaires, les gens en retraite, et leurs familles) dispose du même vocabulaire que le FrL qui a été documenté entre 1920 et 1940. Pourtant, il est question de registre aussi, car bien des Acadiens savent parler créole et s'en servent de préférence car il est plus intéressant, plus chaleureux, ou plus amusant. Abstraction faite du bilinguisme, les Acadiens (bilingues ou non) ne sont pas conscients de créoliser mais en quelques lieux leur langage est fortement créolisé. Je crois pouvoir localiser ces dialectes d'ici peu de temps, à l'aide d'un questionnaire qui vise à obtenir ce qui survit de toutes les reliques et les créolismes déjà documentés, sur les plans lexical et phonologique.

IV. LES LEXIQUES COMPARÉS

D'après le lexique en abrégé, les deux parlers partagent un lexique de fonds qui est composé en grande partie du FrMar et du FrReg. En effet, je n'ai pas encore trouvé de 'maritimismes' dans l'AcL qui n'existe pas dans le CrL, ce qui prouve une stabilité dans ce registre pendant plus qu'un siècle. La présence des mêmes mots dans les créoles mondiaux confirment leur origine dans le FrMar. Ensuite, qui doute la 'pureté' du registre AcL a toujours recours à l'AcM comme groupe de contrôle, et ce dernier n'est pas moins connu pour son vocabulaire maritime.

Le deuxième groupe dans le lexique, les régionalismes, doit dater également de la formation de ces deux parlers. Le point est vérifié pour l'AcL par le corpus AcM. Quant au CrL, il est difficile de juger de la date d'entreé de ces régionalismes dans le lexique: il peut s'agir d'une ré-lexification très poussée sous la pression de l'AcL, mais ces mots ont pu exister dans le français familier de toute personne du dix-huitième siècle. Ensuite, les acadianismes ne sont pas nombreux parce qu'on ne peut pas en signaler beaucoup qui

seraient uniquement de cette source, comme 'ouaouaron,' qui vient des Indiens dans le Nord.

C'est le lexique créole qui n'est pas partagé par l'AcL qui fait voir jusqu'à quel point ces deux parlers sont éloignés. Les verbes de base sont du créole seulement, et en acadien on continue à conjuguer les verbes, donc la morphologie verbale est intacte sauf dans le cas de quelques assimilations du type 'je vas, tu vas' et 'je parlons, nous parlons.' Ces usages sont loin d'être universels, en plus, et montrent encore la fragmentation du dialecte AcL.

Par contre, les créolismes assimilés par l'AcL sont des morphèmes assez importants, à l'ordre des pronoms relatifs ou interrogatifs et même de la morphologie verbale. L'indicateur /té/ remplace 'était' comme auxiliaire du passé composé en quelques dialectes, mais en d'autres il remplace les deux auxiliaires du passé 'être' et 'avoir' (dans toutes les formes). Un autre exemple morphologique important est les formes /mo/ et /to/ pour 'moi' et 'toi,' mots de fréquence élevée dans le discours normal, mais ces formes ont plus à voir avec la phonologie que le lexique.

V. LES SYSTEMES PHONOLOGIQUES COMPARÉS

C'est dans la phonologie justement que les mauvaises identifications des créolismes se font car ils ont plusieurs points communs avec les dialectalismes de l'AcL, en dépit de leurs origines distinctes.

D'abord, un examen des caractères saillants du créole est nécessaire. Le créole imite fidèlement le timbre des voyelles de la langue de contact, mais à l'origine il n'a pas possédé de voyelles centrales, la série labialisée, ce qui a entraîné plusieurs assimilations; le schéma indique les réalisations créoles au-dessus des voyelles d'origine dans le FrM:

CrL	ã/õ	ɛ̃	i	e	ɛ	a	o	u	j
	∧	∧	∧	∧	∧	∧	∧	|	|
FrMar	ã/õ	ɛ̃/œ̃	i y	e ø	ɛ œ	a ɑ	o ɔ	u	j

Les voyelles nasalisées postérieures ont déjà été assimilées dans le FrMar, ce qui est un dialectalisme des dialectes de l'Ouest de la France. L'effet le plus caractéristique du CrL est le refus de labialisation, et on peut voir qu'il est systématique en notant le désarrondissement systématique des voyelles antérieures arrondies du FrMar. Forcément, les voyelles /e/ et /ɛ/ ne sont plus des allophones mais des phonèmes, car elles distinguent l'ouverture dans deux séries de voyelles antérieures médianes dans la langue de source, i.e., les voyelles antérieures arrondies et écartées, soit quatre voyelles de source. Il est vrai que la différence allophonique est conditionnée par la structure de la syllabe en FrSt mais cette structure est parfois annulée par l'amuïssement de /r/ finale et la prononciation des consonnes en finale qui ont dû être muettes à l'époque du contact. La simplification du schéma des voyelles peut poser un problème en comblant trop de distinctions

phonologiques mais le CrL dispose aussi de l'agglutination de l'article défini et de la restauration des consonnes finales pour empêcher des malentendus:

FrMar: paix, peu, père, peur CrL: /lape/, /pe/, /pɛr/, /pɛ/

Autrement, le CrL est composé de nombreuses transformations de la phonologie du français moderne: la métaphonie (i.e., la recherche du mème son dans une syllabe, ce qui entraîne la réduplication des voyelles), les métathèses (apparemment sans motivation), l'affrication systématique des chuintantes palatales /ʒ/ et /ʃ/, et la nasalisation progressive qui ré-interprète la tendance normande de nasaliser /ɛ/ (ouverte) en finale, particulièrement dans les verbes (une caractéristique attestée dans le FrReg même à l'époque moderne).

La deuxième appendice donne un échantillon du CrL, des extraits des études avec traductions des professeurs créolophones cités plus haut.

Un profil de la phonologie acadienne est assez ressemblante au CrL dans les caractères suivants: des métaphonies rares (limitées au mots familiers et au parler enfantin, e.g., /zozo/ 'oiseau' et /piti/ 'petit,' une variante affective qui est entièrement assimilée dans quelques dialectes); des métathèses heritées du FrReg, e.g., /freme/ 'fermer,' /fordoʃ – frdoʃ – fɛrdoʃ – fardoʃ/ 'branchailles, bois mort,' /furmi – fromi/ 'fourmis,' etc.); la conservation des consonnes affriquées à l'initiale dans quelques variantes affectives, e.g., /ʧy/ 'cul, derrière d'une personne,' /ʤab/ 'diable,' /ʤœl/ 'gueule,' et la nasalisation spontanée de l'/ɛ/ ouverte libre en finale, e.g., /ʃɛ̃/ 'cher,' /ɛ̃mɛ̃/ 'aimer.' Ce dernier est systématique dans quelques dialectes et, à l'exception du mot 'cher,' limité aux verbes.

Ce profil phonologique grossier correspond presque exactement à la description du CrL de plus haut, mais ce n'est pas l'AcL qui s'assimile au CrL, ni le contraire: ces quatre caractéristiques datent de la constitution de ces deux langues, car ils trouvent leurs origines dans le FrMar et les dialectalismes du Poitou et de la Normandie, comme le prouvent toujours les atlas linguistiques de Poitou et de France, deux ou trois siècles après la débarcation de ces deux groupes dans le Nouveau Monde.

La persistence de ces caractères en France jusque dans ce siècle donne une idée de leur importance au passé, bien qu'un caveat s'impose dans l'emploi de ces deux sources. L' ALO donne des dialectalismes et l' ALF donne des régionalismes comme attestations, et la distinction relève des différences méthodologiques dans le travail des enquéteurs: l' ALO vise à étudier des reliques pour savoir quelles ont été les formes périmées de ce dialecte à l'époque la plus reculée que possible; donc, Massignon a fait des enquêtes presque uniquement sur des vocabulaires techniques des métiers traditionnels. Effectivement, elle a réussi à sonder un dialecte, là où Gilliéron s'est limité aux notions quotidiennes pour esquisser les différences dialectales du parler français national. Comme résultat, son atlas marche sur des régionalismes, et les atlas linguistiques par régions – tels que celui de Massignon – ont recensé des dialectes, pas seulement des registres variés du FrSt. Encore la question est-elle une question de chronologie mais elle comporte des différences dans les données qui sont susceptibles d'y apparaître.

APPENDICE I

VOCABULAIRE LOUISIANAIS COMMUN AU CRÉOLE (CRL) ET À L'ACADIEN (ACL)

I. 'MARITIMISMES'

bord (côte), Hull
gréyer (préparer), Hull
amarrer (lier), Fortier
larguer, Hull
parer (préparer), Hull
boucane (fumée), Hull
macaque (singe), Hull
ga (regarde[z])
néyer (noyer), Fortier
au ras (dans, près)
tiens-bondre (rester fort)
haler (tirer)
hardes (equipement; vêtements)
butin (possessions)
au large (en terrain ouvert)
frète (froid), Broussard
guetter (regarder), Fortier
héler (appeler)

II. VOCABULAIRE COMMUN PROVENANT DU FRANÇAIS RÉGIONAL (FRR)

vilain (mauvais), Fortier
garli (galerie)
babiller (réprimander)
asteur (maintenant)
icitte
ɡeule (<gueule, bouche)
filer (s'en aller vite)
jongler (?) (penser)
arien (rien)
mander (demander)
zozo (oiseau)
ano (allons)
vaillant (bon, brave, agréable)
tracasser (gêner, embêter), Fortier
aussite
abîmer (battre, < injurier)
maigner (rester)
/moʃe/ (moitié)
yinque (rien que)

III. ACADIANISMES

chassepareille (salseparilla), Massignon
ouaouaron (grenouille géante d'Amérique), Massignon
maringouin

IV. CRÉOLISMES PAS ASSIMILÉS PAR L'ACL

/gɛ̃/ < gain
/kuri/ < courir
/vini/ < venir
/olé/ < voulez
/ka/ < capable
/koté/ < (du) côté (de)
'avoir' (toutes les personnes)
'aller'
'venir'
'vouloir'
'(etre) capable (de)'
'chez'

V. CRÉOLISMES ASSIMILÉS PAR BEAUCOUP D'ACADIENS DE LOUISIANE

/kofè/ < quoi faire?,	'pourquoi?'
/a fɔrs k/ < de force que,	'parce que'
/ki sé/ < qui c'est,	'qu'est-ce que c'est?'
/té/ < était, étaient	(auxiliaire du passé composé, employé avec l'infinitif du verbe)
/apé/ < être après,	'être en train de'
/piti/ < petit,	'petit'

VI. ANGLICISMES LONGTEMPS ASSIMILÉS AUX DEUX

/fun/ < fun,	'divertissement'; e.g., 'gain foune avec' (rire aux dépens de l'autre)
/trɔs/ < trust,	'se fier à'
/bɔs/ < bust,	's'éclater'
/bæd lɔk/ < bad luck,	'mauvaise chance'

APPENDICE II

Extraits de *Louisiana Creole Dialect* (James F. Broussard; Louisiana State University Press, 1942) et de *Louisiana Folk-tales in French Dialect and English Translation* (Alcée Fortier; New Orleans: Hansell, 1894)

POÉSIE DE 'PIERRE'; TRADUCTION DU DR. BROUSSARD

'Cofé?'	'Pourquoi?'
Cofé Bon Djé pas blanchi moin?	'Pourquoi le Bon Dieu ne m'a pas fait blanc?
Mo fait tout ça 'tit Maite li fait.	Je fais tout ce que fait mon maitre.
Mo monte mo choual plis fou que li;	Je monte mon cheval plus hardiment que lui;
Mo ca chanter plis doux que li;	Je chante plus doux que lui;
Mo ca vini plis saoul que li;	Je bois plus saoul que lui:
Mo fait tout ca 'tit Maite li fait.	Je fais tout ce que fait mon maitre.
Cofé bon Djé pas blanchi moin?	Pourquoi le bon Dieu ne m'a pas fait blanc?
Cofé Bon Djé pas blanchi moin?	Pourquoi le Bon Dieu ne ma pas fait blanc?
Mo fait tout ça 'tit Maite li fait.	Je fais tout ce que fait mon maitre.
Mo ca l'aimain plis dir que li;	Je peux aimer plus dur que lui;
Mo ca tirer plis sir que li;	Je tire le fusil plus sur que lui;
Ma ca mouri plis mir que li;	Je meurs plus mur que lui.
Mo fait tout ça 'tit Maite li fait.	Je fais tout ce que fait mon maitre.
Cofé bon Djé pas blanchi moin?	Pourquoi le bon Dieu ne m'a pas fait blanc?

'MARIE MOURI'

Pitit zozo, ça t'apé fait?
T'apé sauter, t'apé chanter?
To pas connain n'a plis Marie?
Marie mouri, Marie mouri!

Tits zerb' to verts, tits zerb' to mou,
Faut plis to fait en lite pou' nous.
To pas connain n'a plis Marie?
Marie mouri, Marie mouri!

Quand jou' vini, n'a pas soleil
La nouitte vini, n'a pas sommeil.
Quand moune content, mo pas ca ri
Marie mouri, Marie mouri!

'MARIE EST MORTE'

Petit oiseau, que fais-tu?
Tu sautes, tu chantes?
Tu ne sais pas que Marie n'est plus?
Marie est morte, Marie est morte!

Mon petit gazon, tu es vert, tu es mou
Il ne faut plus que tu fasses un lit pour nous.
Tu ne sais pas que Marie n'est plus?
Marie est morte, Marie est morte!

Quand vient le jour, il n'y a pas de soleil,
Quand vient la nuit, il n'y a pas de sommeil.
Je ne peux plus rire avec les heureux,
Marie est morte, Marie est morte!

EXTRAIT DU 'MARIAZE DE COMPERE LAPIN' (FORTIER, 1894)

–Ou tap couri comme ça, galopé,galopé tout temps?
–Ah, réponde Compair Lapin, vous pas connin mauvais nouvelle que Lion déclare la djerre tous nélé-phants et ma pé verti tous milets, choals, et chameaux yé pou ye fou
–Mais to même qui zaffaire to gagso? nin pou galopé, yé pas apé prend toi pou fait soldat avec toi?
–Non, to crois ça, réponde Compair Lapin, ah bien, to pas connin arien avec tout to malin.
Quand n'officier le roi a vini chercher choals et milets pou la cavalerie pou fait la djerre yé va dit comme ça:
Alà ein bougue grand zoreille, c'est ein milet et mo sra fouti, yé va enrolé moin et moi va bligé marché. Mais semble moin mo connin vous, mais si longtemps mo pas oua vous. Bon Djé tende moin, c'est Renard, mo zami lezotes fois!
–Oui, oui, c'est moin, mo vié. Eh bèn, ça vous dit pou tout vilain affaire yé?
–Tout ça pou ein femme, dit Compair Lapin, faut nous séyé, mo zami, pas trouve nouzottes dans yé procès.

–Where are you going like that, running all the time?
–Ah, replied Compair Lapin, you don't know the bad news. Lion has de-clared war against all elephants, and I want to notify all mules, horses, and camels to get out of the way.
– But you, why are you running? They are surely not going make a soldier out of you?
–No, you believe that. Ah, well, all your cunning, you know nothing.
When the officers of the king will come to get the horses and mules for the cavalry to go to war, they will say: That's a fellow with long ears; he is a mule; let's take him. I will be caught and will have to march. But it seems to me that I know you, but it seems such a long time since I've seen you. God help me, it's Master Fox, my old friend!
–Yes, yes, it is I, my old friend. Well, what do you say about all this ruckus?
–All that for a woman! We must try to have nothing to do with it.

REFERENCES

Broussard, James (1942). *Louisiana Creole Dialect.* Baton Rouge: Louisiana State University Press.

Conrad, Glenn (1975). 'The Culture of Acadiana: An Anthropological Perspective,' in *Tradition and Change in Southern Louisiana,* ed. Steven L. Del Sesto and Jon L. Gibson. Lafayette: University of Southern Louisiana.

Conwell, Marilyn J. et Alphonse Juilland (1963). *Louisiana French Grammar.* The Hague: Mouton.

Fortier, Alcée. 1894. *Louisiana Studies: Literature, Customs and Dialects, History and Education.* New Orleans: F. F. Hansell & Sons.

Hull, Alexander (1968). 'The Origins of New World French Phonology,' in *Linguistic Studies Presented to Andre Martinet on the Occasion of His Sixty-eighth Birthday,* Pt. II: Indo-European Linguistics, (ed. Alphonse Juilland). New York: Linguistic Circle of New York, 1968 [appeared 1970]. [also published in *Word* 24 (for 1968)].

Massignon, Geneviève. 1962. *Les Parlers français d'Acadie.* Paris: Klincksieck.

Morgan, Raleigh, Jr. 'The Lexicon of St. Martinville Creole', *Annales Linguistiques,* II, i, 7-29.

Robin, C. C. 1804. *Voyage à l'intérieur de la Louisiane.* Paris.

Trappey, Maud 1940. *The French Speech of Iberia Parish.* Thesis, Louisiana State University.

STRUCTURAL INFLUENCE OF ARABIC AND PERSIAN ON THE NORTH-WESTERN INDO-ARYAN LANGUAGES

Vit Bubenik
Memorial University of Newfoundland

ABSTRACT

The structural influence of Classical Arabic and literary/spoken Persian on Hindi-Urdu and other Indo-Aryan (IA) languages falls into the category of slight structural borrowing. Stronger structural influence is to be expected in the westernmost IA languages, Sindhi and Lahnda, spoken in the areas where the natives were exposed to spoken Arabic already in the 8th c. They are unique among the other IA languages in attaching pronominal clitics to nouns, verbs and in colloquial Sindhi even to post-positions. Some effort will be made to sort out several factors: Dardic substrate, grammatical borrowing from Arabic and Persian, pan-IA innovations and the influence of Hindi-Urdu.

1. INTRODUCTION

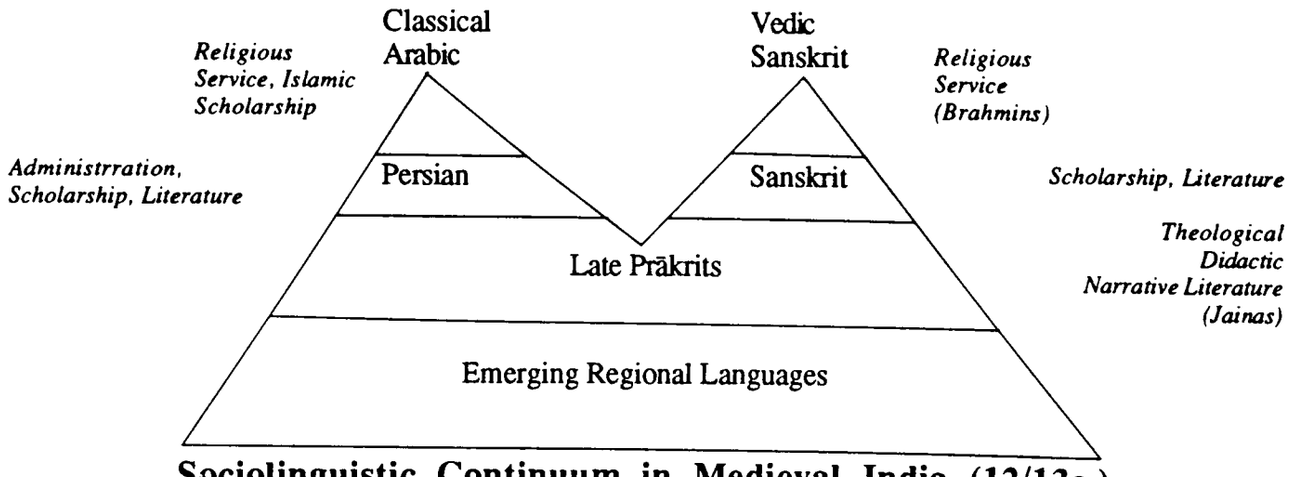
The structural influence of Classical Arabic on the languages of various Moslem peoples falls into the category of slight structural borrowing (to use Thomason and Kaufman's borrowing scale 1988:74-6). In the case of North India we are dealing with a prestigious language which was used by the Moslem invaders – ethnic Turks, Persians and Afghans – for literary and religious purposes. Its influence on the late Middle Indo-Aryan languages – starting in the 11th c. in the Gangetic Doab but several centuries earlier in Sind and Punjab – was mediated through three channels. Only the first one involved actual oral bilingualism among speakers of late Prākrits in Sind and Punjab in the 8th c. onwards; the other one involved the Persian intermediary and the third one – without Arabic-Prākrits or Persian-Prākrit oral bilingualism – would be limited to situations when Hindu converts to Islam could hear Classical Arabic recited in religious services. A propos the Persian intermediary, Persian became the administrative language of the Delhi sultanate (1200-1526) and the subsequent Moghul empire until 1835. (It is somewhat astonishing to realize that the literary output of India in Persian is bigger than that of Iran itself). Persian – in its turn – underwent a structural interference from Arabic several centuries ago, starting with the conquest of Iran at the end of the 7th c. Its details are beyond the scope of this paper; broadly speaking, we may assume a process of merging by which the Arab invaders and later settlers were slowly iranicized (cf. Versteegh 1984:66). After a few centuries Middle Persian (i.e., Parsi, later called Pahlavi) was reinstated as the national language but Arabic retained its position as the language of religion, science and in the beginning, of the administration.

The historical events which brought Arabic and Persian to India cannot be recapitulated here. Most succintly, the Indian subcontinent during the 10th c. was divided into a number of kingdoms ruled by native dynasties whose power was exhausted by internal and external fights. Then in the year 997 a Turkish nobleman, Mahmud of Ghazni, moved for the first time his army from the mountains of the Hindu Kush through the Khyber Pass to the Punjab plains. Until 1030 these campaigns became almost an annual event. The second attack came under the leadership of Muhammad Ghuri at the end of the

12th c. He entered the Indus plain through the more southerly Gomal Pass and a series of campaigns followed: in 1185 Muhammad conquered Lahore, and in 1192 he defeated the last king of Delhi, Prthviraja, and his kingdom fell to Muhammad. The latter was assassinated in 1206 but – unlike the previous occasions – this did not mean the temporary withdrawal of the Turkish-Afghan army from India. His general, Qutb-ud-din Aibak, became the ruler of a new political entity in India – the Delhi sultanate which lasted until 1526.

Let us recall an earlier encounter of India with ethnic Arabs which took place at the beginning of the 8th c. In 712 the westernmost part of India, namely the lower valley of Indus called Sind, became the easternmost extremity of the general Arab expansion through Asia. In Sind the Arabs under Muhammad bin Qasim did not meet with any major resistance since the area was largely desert, but their attempts at further conquests northwards and eastwards were resisted: for instance, we know from the Annals of the Tang dynasty that Candrapida, the king of Kashmir, asked for assistance from the Chinese emperor to repel Arab attacks, and another king, Lalitaditya (724-61), was successful in pushing back the Arab forces in the Punjab (Kapur 1983:29). It is of fundamental importance for our further investigation to realize that the Indo-Aryan inhabitants of the Indus valley met ethnic Arabs and were exposed to spoken Arabic. Given the time limits, we cannot enter into the discussion of the pattern of Arab settlement which resulted in intermarriages, Arabic-Sindhi bilingualism and rapid islamization. Suffice it to mention that the Koran was translated into Old Sindhi in 883 and that there was a Persian version of the epos Mahabharata translated from Arabic in 1026 (the latter in its turn was translated from Old Sindhi as spoken ca. 1000, cf. Khubchandandi 1969:215). Summarily, the regions in the lower and middle parts of the Indus valley were linguistically Arabo-Persianized and culturally islamized several centuries before similar processes started taking place in the heart of Northern India, i.e., the Gangetic Doab. Here the influence of Classical Arabic was channeled through spoken Persian or through religious services and high literature composed by ethnic Afghans and Turks (starting in the 12/13th c.).

As a result of all these complicated population movements and demographic changes the upper social strata of medieval India were divided culturally and linguistically as shown in Diagram 1. Classical Arabic – as a language of religious service and Islamic scholarship – and Persian – as a language of administration and literature (esp. poetry) – appeared on the top of the sociolinguistic continuum to match Vedic and Classical Sanskrit used by Hindus in parallel functions.



Sociolinguistic Continuum in Medieval India (12/13c.)

2. SOME CHARACTERISTICS OF THE NORTH-WESTERN GROUP OF INDO-ARYAN LANGUAGES

At this point I want to introduce the two North-West I-A languages whose borrowings we want to examine in some detail, namely Sindhi and Lahnda. Sindhi is spoken in the southernmost province of Pakistan in three regional varieties called Vicōṯ, Siraikī and Lāṛī. Vicōṯ, spoken in the city of Hyderabad and its surroundings, is considered standard Sindhi. North of Sindhi lies the troubled area of so-called Lahnda, divided usually into Siraiki and 'Northern Lahnda' (cf. Shackle 1976:1-2). The latter variety is known also misleadingly by its geographic label 'Western Panjabi' (cf. Grierson 1919:233 ff.). Siraiki is spoken in the environments of the city of Multan (hence the label Multani). In its northern section Lahnda is surrounded by languages of the Dardic type on three sides: Kashmiri, Shina and Kohistani. Both Sindhi and Lahnda lie on the western edge of the I-A language area, their western neighbors being the East Iranian languages Balochi and Pashto. And finally, all the Pakistani languages are overlaid by supraregional Urdu, the official language of administration and education in Pakistan, which also serves as the chief vehicle for all types of writing. Both Sindhi and Lahnda possess a small body of literature; e.g., recently, there has been a great expansion in the use of Siraiki in the Multan-Bahawalpur area. As far as Sindhi is concerned, the above mentioned Old Sindhi translations of Koran and Mahabharata did not come down to us and the extant Sindhi literature goes back only to the 17th c. The principal literary language of the area under consideration was Persian until its replacement in the 19th c. by Urdu in the Panjab and by Sindhi in Sind.

Sindhi and Lahnda form together the North-Western group of I-A languages. This group possesses many characteristics which connect it with the Dardic languages, a big and heterogeneous group of ca. 15 languages, including Kashmiri and Shina. Among the isoglosses which separate the North-Western group from the rest of the I-A languages but which link them with the Dardic group we may mention the preservation of the intervocalic voiceless dental stop, certain pronominal forms (e.g., Lahnda *assī* 'we', *tussī* 'ye'), the use of pronominal clitics to finitize past participles and the archaic ergative construction continuing the MIA antiabsolutive (Bubenik 1989).

3. PHONOLOGICAL AND MORPHOLOGICAL BORROWINGS IN SINDHI AND HINDI-URDU

Slight structural borrowing includes phonological borrowing usually confined to loan-words. Here we may start with the well-known case of the phonological system of Urdu which contains three phonemes borrowed from Arabic and Persian: voiceless labiodental fricative /f/, voiced dental fricative /z/ and voiceless velar fricative /x/. These three are found in the speech of educated Moslems, whereas other Moslems and Hindus replace /f/ and /x/ by native voiceless aspirates /ph/ and /kh/, and /z/ by the palatal affricate /ǰ/. Sindhi, in addition to /f/, /z/ and /x/, borrowed also the voiceless uvular stop /q/ and adapted the voiced uvular fricative /ġ/ as the voiced velar fricative /ɣ/. The latter two phonemes, however, may also be alternatively realized as plain velar stops [k] and [g] (cf. Egorova 1966:20). In Hindi these two phonemes were completely nativized as plain velar stops. In Sindhi and Urdu they are spelled differently from the plain counterparts in the Arabic script, but even in writing Hindi-Urdu in Devanagari it is a sign of good education to dot the letters K and G. Some representative examples are displayed in (1) showing the

Arabic source words and their shape in Sindhi and Hindi-Urdu, the latter subdivided into educated Moslem vs. other Moslems and Hindus.

(1) **Selected Arabic loan-words in Sindhi and Hindi-Urdu**

Arabic	**Sindhi**	**Hindi-Urdu**		
source (8th c.)	Moslems	Moslems (educated)	Moslems and Hindus	
fikru	fikru	fikr	phikr	'thought'
manzilu	manzilu	manzil	mãǰil	'storey'
xatmu	xatmu	xatm	khatm	'end'
qalamu	qalamu ~kalamu	qalam	kalam	'pen'
ġarību	ɣarību ~garību	ɣarīb	garīb	'poor'

As far as the final *-u* in Arabic and Sindhi are concerned, *-u* in Arabic is the suffix of the nominative which survived from Classical Arabic into the 8th c. (re *iʕrāb* cf. Versteegh 1984:4-5); the final *-u* in Sindhi, realized as lax [ʊ] in English, is a remarkable archaism from the late Middle Indo-Aryan stage: Sindhi is the only I-A language which preserved the Apabhraṃsa suffix of the absolutive case in *-u*.

In derivational morphology, the affixes may be abstracted from borrowed words and added to native vocabulary. For instance, to express concepts corresponding to English adjectival derivatives formed by the privative suffix *-less*, both Sindhi and Hindi-Urdu may use either the Arabic prefix *lā-* or the Persian prefix *be-* or the native Sanskritic prefix *nis/r-*; several examples are displayed in (2).

(2) **Adjectival derivatives of Sindhi and Hindi-Urdu with Arabic and Persian prefixes.**

	Sindhi	**Urdu**	**Hindi**
'childless'	beawlādu	lāvalad	nissātān, nirvāś (*putrarahit, putrahīn* in High Hindi)
'speechless'	lājavābu	lājavāb	niruttar
'careless'	befikru	befikr	niśčītit

Urdu *lāvalad* 'childless' is in keeping with the major derivational pattern of Classical Arabic for expressing concepts such as the English adjectival derivatives with *anti-* or *-less* (e.g., *lā sāmiyya* 'anti-semitism', *lā anāniyya* 'selflessness', etc.). The Sindhi derivative *beawlādu*, however, is a hybrid combining the Persian prefix be- with the Arabic

plural form of the word *walad* 'child'. The plural derivational base is used only marginally in Arabic, as in one of the two words for 'republic': *ǰumh+riyya* vs. ǰamāhiriyya (the latter one being the official title of the Republic of Libya):

(3)

	'childless'		cf. Arabic derivational patterns	
Urdu	lāvalad not-child	'childless'	ǰumhūr-iyya mass (of people) - abstract suffix	'republic'
Sindhi	beawlādu without children	'childless'	ǰamāhir-iyya masses (of people) - abstract suffix	'republic'

In inflectional morphology, foreign affixes may enter the borrowing language but remain confined to borrowed vocabulary items. Both Sindhi and Urdu borrowed the Persian genitival construction, so-called *ezafet*, whose NG and NA orders are not in harmony with their overall SOV typology. Several examples are given in (4).

(4) **Persian genitival construction in Sindhi and Urdu (limited to Arabic and Persian borrowings).**

	Sindhi	Urdu	Hindi
'Prime Minister"	vazīr-e-sarkār	vazīr-e-āzam (~ājam)	pradhān mātrī (< Sanskrit)
'(hall of) public audience'	darbār-e-ām	darbār-e-ām	
'Moslem's arm's strength'		kūwat-e-bāzu-e-muslim (Iqbāl)	

Both Sindhi and Urdu borrowed the morphology of the comparative and superlative from Persian; the latter one is limited to two lexical items, namely 'worst' and 'best' (*badtarīn* and *behtarīn)*: the former shows some limited productivity in High Hindi since the Persian suffix *-tar* happens to be identical with the Sanskritic suffix *-tar*. High Hindi may also use the Sanskritic superlative suffix *-tam*; both suffixes, however, are limited to the adjectives borrowed from Sanskrit and they are largely confined to the written language (cf. McGregor 1977:93). These examples are given in (5):

(5) **Comparative & Superlative in Persian, Sindhi, High Urdu and High Hindi**

	Persian		Sindhi, High Urdu	High Hindi
'better'	behtar	--->	behtar	śreṣṭhatar uccatar
'best'	behtarīn	--->	behtarīn	sarvottam

In coordinate compounds Sindhi keeps the Persian conjunction *o* 'and' to a greater extent than Urdu. This is shown in (6):

(6) **Coordinate compounds in Sindhi and Urdu.**

	Sindhi	Urdu
'commerce'	xarīd-o-faroxat	xarīd-faroxt
	purchase-and-sale	purchase-sale

4. MORPHOSYNTACTIC BORROWINGS IN SINDHI AND LAHNDA

In this section several cases of more severe structural borrowing involving morphosyntactic patterns will be discussed. Let us begin with a sweeping typological statement that Sindhi and Lahnda – not unlike all other I-A languages – are basically of SOV GN AN typology. Upon closer examination this statement will have to be considerably qualified in at least five respects:

i) Colloquial Sindhi – unlike rigid SOV languages – may place adverbial phrases after the verb (the same phenomenon is known from colloquial Persian). This is shown in (7).

(7) **Position of the adverbial phrase in colloquial Sindhi vs. Hindi-Urdu.**

'the husband went to Africa'

Sindhi	patī husband	halyo went	vyo AUX	āfrīkā Africa
cf. **Persian** (colloquial)	šauhar husband	raft went	be-āfrīkā to Africa	
vs. **Hindi-Urdu**	pati husband	afrīkā Africa	calā went	gayā AUX

(ii) More seriously, in the formation of the present tense by means of the auxiliary *tho* Sindhi may place it not only after but also before the main verb. Again, in Hindi-Urdu – as in all other I-A languages – the auxiliaries may appear only in postverbal position.

Persian, however, allows for both positions (more specifically, the copula in the perfective forms appears after the main verb, but the auxiliary of the future tense has to be placed before; contrast *porsīde=am* 'I have asked' with *xvāham porsīd* 'I will ask'). In India only Dardic Kashmiri (of SVO typology) allows for the auxiliaries to be placed before the main verb, e.g., *bI chus go:mut* (I AUX went) 'I went'. These matters are surveyed in (8).

(8) **Verb Phrase in Sindhi vs Hindi-Urdu**

	V=AUX	AUX V
Sindhi (SOV)	halā=tho 'I go'	=tho halā 'I go'
cf. **Persian** (SOV)	porsīde=am 'I have asked'	xvāham porsīd 'I will ask'
cf. **Kashmiri** (SVO)	—	chus go:mut 'I went'
vs. **Hindi-Urdu** (SOV) + all other I-A Ls	caltā thā 'I went'	—

iii) Even more seriously, Sindhi and Lahnda are unique among I-A languages in attaching their pronominal clitics to nouns in violation of their basic GN AN typology. In contemporary Sindhi only nouns denoting relationships and parts of body (i.e., those expressing inalienable possession) may be provided with pronominal suffixes. Furthermore, the suffixes of the 1st and 2nd Pl are not used any more (but they were common in the 19th c., cf. Khubchandani 1962:77). As far as Lahnda is concerned, the pronominal suffixes in all persons were common in the 19th c. (cf. Wilson 1899); more recently, however, Shackle (1976:101) reported that in Siraiki the suffixes are not attached to the nouns any more. Some data are shown in (9):

(9) **Possessive Clitics in Sindhi and Lahnda.**

Sindhi (contemporary)		
ḍ̃āḍ̃u=me 'my grandfather' grandfather=me	~	mãhẽjo ḍ̃āḍ̃o
ḍ̃āḍ̃u=se 'his grandfather'	~	hu/anajo ḍ̃āḍ̃o

Lahnda (Shahpur, 19th c.)	**Lahnda** (contemporary)
ghar=am 'my home' home=me	merā ghar
ghar=us 'his home' home=his	us/hdā ghar

iv) In contemporary informal Sindhi – but not in Lahnda – even the postpositions may host the pronominal clitics; in a sense, the postpositions behave here like prepositions. In its formal code, Sindhi – as all the other I-A languages – uses postpositions after the oblique pronominal forms. This is shown in (10).

(10) **'Postpositions' as phonological hosts to pronominal clitics in Sindhi.**

	formal code (oblique Pro=Po)	*informal code* (Pr=Cl)
'with me'	mū-sāṇu	sāṇu=me
'with you'	to=sāṇu	sāṇu=(h)e
'with him'	huna=sāṇu	sāṇu=se
'to him'	huna=khe	khe=se

v) And finally, both Sindhi and Lahnda use pronominal clitics with the past participle to encode the pronominal agent and the goal. Consider a simple sentence 'I saw him' which has to be expressed by the ergative construction in all western I-A languages. As shown in (11) the only difference between (formal) Sindhi and Hindi-Urdu will be the absence of the ergative postposition on the oblique pronominal form in the former language:

(11)

mũ	huna=khe	ḍḍiṭho	'I saw him' (formal Sindhi)
I+OBL	he+OBL=DAT/ACC	see+PP+M	
mɛ̃=ne	us=ko	dekhā	(Hindi-Urdu)
I=ERG	he+OBL=DAT/ACC	see+PP+M	

Informal Sindhi, however, may cliticize the pronominal agent to the past participle, and even the pronominal goal on top of it, as shown in (12):

(12)

khe=s	ḍḍiṭhu+m	'I saw him'
DAT/ACC=3/Sg	see+PP+1/Sg	(informal Sindhi, with enclisis of Ag)
ḍḍiṭho+mã+s		'I saw him'
see+PP+1/Sg+3/Sg		(informal Sindhi, with synenclisis of Ag and Go)

The first strategy – called antiabsolutive by Hook (1984) – is known also from Kashmiri, cf. *wuchu=m*, which means literally '[he is] my seen'. There are also Kashmiri parallels to the synenclisis of pronominal agent and goal; and furthermore, this type of synenclisis is fairly common in colloquial Persian, and the only means available in Arabic. These examples are given in (13):

(13) wuchu=m 'I saw him' (Kashmiri 'antiabsolutive')
see+PP+1/Sg
lit. [he is] my seen

dīd+am+eš 'I saw him' (synenclisis in colloquial Persian)
saw+1/Sg+3/Sg

ra?ay+tu+hū 'I saw him' (synenclisis in Classical Arabic)
saw+1/Sg+3/Sg

Diachronically speaking, we possess some indirect MIA evidence for the enclisis of pronominal agent to the past participle but I am unaware of any evidence for the synenclisis of Ag and Go and/or Rec; similarly, the nouns could host pronominal clitics (especially, if found in S-1 position); however, there is no evidence for the postpositions hosting clitics. We may thus cautiously formulate our conclusions in a sense that the North-West I-A languages in their usage of pronominal clitics with nouns and verbs have continued previously existing MIA structures; however, their retention and further development (vis-à-vis all other I-A languages) is best explained by structural interference from Arabic and Persian. This assumption is further confirmed by the use of clitics with postpositions in Sindhi which is undocumented in MIA. There is an interesting parallel to the borrowing of pronominal enclisis from the same geographic area. Brahui – a Dravidian pocket language surrounded by Balochi – is the only Dravidian language which developed (from native morphemes) pronominal clitics used with nouns as possessive suffixes and as the goal with verbs. Emeneau (1962:60) explained the Brahui pattern as coming either from Iranian or from the North-West I-A. It is of interest to observe that contemporary Balochi – like Lahnda mentioned above – stopped using the possessive clitics with nouns; presumably, under the influence of Urdu, and Urdu and Panjabi in the latter case.

5. CONCLUSIONS

Thomason and Kaufman (1988:77-95) considered a number of cases from the extraordinary complex Indian *Sprachbund* and assigned them numerical values on their borrowing scale from (1) to (5). The value (2), 'slight structural borrowing', has been assigned to the influence of Classical Arabic on Hindi-Urdu; (3), 'slightly more structural borrowing', to the Sanskritization of literary Dravidian languages, and (4) 'moderate structural borrowing', to the Indo-Aryanization of non-literary Dravidian languages, and the influence of Balochi on Brahui. For the reasons presented above, the Arabo-Persian influence on the North-West I-A languages could be tentatively placed somewhere between their categories of slight (2) and moderate structural borrowing (4).

REFERENCES

Bubenik, V. 1989. 'An interpretation of split ergativity in Indo-Iranian languages', *Diachronica* VI:2, 181-212.

Egorova, R.P. 1966. *Jazyk Sindhi*. Moskva: Nauka.

Emeneau, M.B. 1962. *Brahui and Dravidian Comparative Grammar*. Berkeley and Los Angeles: University of California Press.

Grierson, G.A. 1968/1919. *Linguistic Survey of India.* Vol. VIII, Parts I and II. Delhi: Motilal Banarsidass.

Kapur, M.L. 1983. *Kingdom of Kashmir.* Jammu: Kashmir History Publications.

Khubchandani, L. 1962. 'Pronominal suffixes in Sindhi', *Indian Linguistics* 23, 72-81.

----------. 1969. 'Sindhi' in T.A. Sebeok (ed.), *Current Trends in Linguistics* Vol. 5, 201-234, The Hague: Mouton.

McGregor, R.S. 1977. *Outline of Hindi Grammar.* Delhi: Oxford University Press.

Shackle, C. 1976. *The Siraiki Language of Central Pakistan.* London: Unwin.

Thomason, S.G. and T. Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics.* Berkeley: University of California Press.

Versteegh, K. 1984. *Pidginization and Creolization: The Case of Arabic.* Amsterdam: Benjamins.

Wilson, J. 1899. *Grammar and Dictionary of Western Panjabi, as spoken in the Shahpur District with Proverbs, Sayings and Verses.* Lahore.

MICMAC NUMERALS: AN ETYMOLOGICAL STUDY

Audrey Dawe-Sheppard
Memorial University of Newfoundland

ABSTRACT

This paper presents an etymological explanation of the Micmac numerals. The Micmac numerals are examined to discern whether they are reflexes of the possible reconstructions labeled by Siebert as the quinary system, the decimal system and the descriptive system. It will be shown that Micmac shows a blending of the three systems. Examples from other Algonkian languages (including both Eastern and Central) will be examined to show that other languages, like Micmac, also have a blending of the systems. The numerals from 1 to 10 will be the major focus; but other numerals which display a noteworthy etymology will be presented and examined.

According to Siebert's 1975 article 'Resurrecting Virginia Algonquian From the Dead', there are three main systems of numerals in the Algonkian languages: 1) the quinary system; 2) the decimal system; and 3) the descriptive system. The quinary and decimal systems are based on the fact that 'In the Algonquian languages the cardinal numerals are based on a primeval digital enumeration in which the hands were customarily held in pronation.' (Siebert 1975:303) [The cognitive reasoning behind the quinary and decimal systems is that a person has two hands with five fingers on each which gives either a base of five or a base of ten.]

Siebert discusses these three systems in some detail. In the quinary system the numbers from 1 to 10 are separated into two groups. The lower numerals from 1 to 5 are counted on the first hand and are formed from elementary roots. The higher numerals from 6 to 10 are counted on the second hand and are formed by compounding the roots of the lower numerals. In the decimal system there is no clear-cut distinction between the lower numerals and the higher numerals. In the Algonkian languages both these systems are often replaced by substitute descriptive sets as will be seen in some of the Micmac data.

In this same article, Siebert proposes reconstructions for the Proto-Algonkian numerals and he fits them into the above categories for the numeral systems. We shall examine these and compare the Micmac numerals to discern whether they are reflexes of these proposed reconstructions. If such is the case, the etymology will show the category or categories the Micmac numerals belong to. It will be seen that Micmac shows a blending of the three systems.

The following chart displays the Micmac numerals from 1 to 10 and Siebert's proposed reconstructions. The corresponding proto-forms for the Micmac numerals are those which are underlined.

Siebert's Proto-Algonkian Reconstructions for the Numerals

Number	Micmac	Quinary	Decimal	Comments
1	newt	*nekwetwi		
1 ('only')	paslk		*pe:syekwi	
2	ta'pu	*nyi:swi	*ta:pawi	
3	si'st	*ne?θwi		Origin not clear
3	nes	*ne?θwi		
4	ne'w	*sa:ka	*nye:wwi	
5	na'n	*nya:θanwi	*nya:nanwi	
6	asukom	*nekwetwa:syeka		Descriptive
7	l'uiknek	*nyi:swa:syeka		Descriptive
8	ukumuljin	*ne?θwa:syeka		Descriptive
9	peskunatek	*sa:ka:syeka	*no:lyiwi	Descriptive
10	mtln	*metaθanθwi (PEA)	*meta:hθwi	
10	newtiska'q	*metaθanθwi (PEA)	*meta:hθwi	

Let us examine each Micmac numeral separately.

1 in Micmac is *newt*. This form is derived from PA *nekwetwi which Siebert has categorized as belonging to the quinary system.

It should be noted that there is also a form in Micmac which appears to be a reflex of *pe:syekwi which is the reconstructed form from the decimal system. Micmac has the form *pasik* which means 'only.'

2 in Micmac is *ta'pu*. This is a reflex of PA *ta:pawi which belongs to the decimal system. PA *ta:paw- is a root meaning 'alike, equal, identical'.

3 in Micmac is *si'st*. Siebert has only reconstructed one form for PA and that is *ne?θwi and *si'st* is obviously not a reflex of this form. Neither does the form appear to be descriptive. Actually, the etymology for *si'st* is unclear.

But there is also an alternate form *nes* meaning 'three' which is derived from PA *ne?θwi as *?θ > s in Micmac. Siebert classifies *ne?θwi as quinary. There are cognates in other Eastern Algonkian languages such as Abenaki *nass* and Penobscot *na`s* (` = preaspiration).

4 in Micmac is *ne'w* which is derived from PA *nye:wwi which Siebert has categorized as belonging to the decimal system. There are cognates of this form in most of the Central Algonkian languages such as Montagnais and Menominee *new*. According to Siebert PA *ny has two different reflexes in Proto-Eastern Algonkian. In some of the languages *ny is reduced to /n-/ as in Micmac *ne'w* and Maliseet *nÉ:o*. In other Eastern Algonkian languages *ny is reduced to /y-/ as in Penobscot *iew*, Natick *yaw* and Abenaki *ieu*. (Siebert 1975:307)

5 in Micmac is *na'n* which could possibly be derived from either of the PA forms *nya:nanwi or *nya:θanwi. It is difficult to discern which of Siebert's reconstructions this

is a reflex of as both would result in *na'n* in Micmac. Therefore it is difficult to categorize 5. It can be belong to either the quinary or the decimal system.

6 in Micmac is *asukom*. This is not a reflex of either of the two forms reconstructed by Siebert: *asukom* does not fit into either the quinary or the decimal system. This is an example of a numeral which is descriptive. Hewson states in his 1987 article 'Micmac Numerals in Betsiamites' that this form clearly contains the Proto-Algonkian root *a:saw- 'across.' (Hewson 1987:18) There are reflexes of the PA root *a:saw in the Algonkian daughter languages as can be seen in the following examples: Menominee and Ojibwa *a:sawaka:m* 'at or on the opposite side' (Hewson in press: 7) which are cognates of Micmac *asukom* and Montagnais *a:shukan* 'bridge'. This indicates a transfer to the opposite hand after the first five fingers of the first hand have been used. The first finger across appears to be the thumb as will be shown by the etymology of 7 which follows. So the form *asukom* indicates 5+1 or the start of a 'new' hand.

7 in Micmac is *l'uiknek* which again is descriptive. It is derivable from the PA form *eθwahikani meaning 'pointer, forefinger' reconstructed by Siebert. This indicates that the first finger across is the thumb, the forefinger being the second. Maliseet and Passamaquoddy also have cognates for this form.

8 in Micmac is *ukumuljin* which is not derivable from the PA forms presented by Siebert. This word can be broken down to give a meaning which indicates that this number is descriptive. *ukum-* is a root which appears to have several meanings one of which is possibly 'tipped over or lopsided' and *-lj-* is a medial meaning 'hand, fingers'. Another possible meaning for *ukum-* is 'lopsided' which could indicate that there is a shift in the weight of the hand. If one is using the hands to count it will be noted that there are more fingers used than not used which gives a lopsided effect. It may also possibly have the meaning 'the finger that sticks out' as the eighth finger does protrude further out than the others.

9 in Micmac is *peskunatek* and this too possibly belongs to the descriptive system. It is reported that it has a meaning 'one finger left' in Maliseet. *Peskunatek* has cognates in the other Eastern Algonkian languages such as Maliseet, Passamaquoddy, Natick, and Narragansett. There are no cognates in the Central Algonkian languages indicating that this is an Eastern innovation.

10 in Micmac is *mtln*. Siebert (1975:305) presents *metaθanθwi as a Proto-Eastern Algonkian reconstruction which he categorizes as belonging to the quinary system. It doesn't really fit into the quinary system as it does not contain a reflex of the PA complex final **a:syeka* - that is, it does not contain the normal compounding element which is used to form the higher numbers (6 to 10) from the lower numbers (1 to 5) in the quinary system. Micmac *mtln* is a reflex of this reconstructed proto-form and there are cognates of this in some of the other Eastern Algonkian languages such as Penobscot.

Micmac also has a variant form for 10 which is *newtiska'q* which is difficult to place into either of the three systems. It should be noted that it does contain the *newt* element which means 'one'. It is not derivable from the forms which Siebert has reconstructed. There is a cognate for this word in Passamaquoddy but seemingly nowhere else in the other Algonkian daughter languages.

From the above discussion it can be seen that Micmac does in fact have a blending of numeral systems. 1, 3 and possibly 5 and 10 are reflexes of what Siebert reconstructs as members of the quinary system but from this it does not necessarily follow that Micmac utilizes a quinary system. As stated earlier the main feature of a true quinary system is that

the higher numerals are formed by compounding the roots of the lower numerals. If a quinary system did exist in Micmac then the numbers 6, 9, and 10 would be formed from 1, 3, and 5 (which according to Siebert's reconstructions are quinary) by the addition of a compounding element. By examining the data it will be seen that this is not the case. Therefore, the evidence for a quinary system in Micmac is non-existent. But there is still a blending of decimal and descriptive.

We may conclude, therefore that 2, 4, and possibly 5 belong to the decimal system; while 6, 7, 8 and 9 are most likely purely descriptive. It is difficult to discern whether the descriptive items are innovations of Micmac or reflexes of an older system. It should be noted that the roots of the words can be traced back to Proto-Algonkian but it is not necessarily the case that these roots were used to form the names for the numerals as they are in Micmac. It seems likely that the use of descriptive items is an Eastern innovation using already existing elements.

There may be something noteworthy here in that Micmac seems to rely most heavily on the descriptive numerals. It has been noted that *neqdensk*, 10 in Passamaquoddy which is the cognate of the variant Micmac form *newtiska'q* means 'one set of fingers' which is descriptive. It has not yet been possible to find this to be the meaning in Micmac but if it is the case then Micmac uses the descriptive system for the entire set of numbers for the second hand; that is the entire set of higher numbers.

[*Newtiska'q* means 'one set of tens' which places it in the decimal system. Another indication of this is that 20, in Micmac, is *tapuiska'q* meaning 'two sets of tens' and 30 *nesiska'q* meaning 'three sets of tens', etc.]

A brief look at a representative language from the Central and Eastern Algonkian groups will show that other languages, like Micmac, also have a blending of systems. The representative language from the Central Algonkian languages is Ottawa which is a dialect of Ojibwa.

Ojibwa Numerals (Masthay 1977)

Number	Ojibwa	Numeral System
1	ningotchua	Quinary
2	nindjwa	Quinary
3	niswa	Quinary
4	niwin	Decimal
5	nanan	Decimal
6	ningotwasswi	Quinary
7	ninjwasswi	Quinary
8	nichwasswi	Quinary
9	shang	?
10	kwetch	?

There is a blending of the quinary and the decimal systems. Ottawa clearly shows how a quinary system works: as can be seen in the table above, the higher numerals 6, 7, and 8 are formed by using the root of the lower corresponding numerals and compounding them with the element *-wasswi*. 4 and 5 have been categorized as decimal based on the reconstructions given by Siebert. It is not possible to categorize 9 and 10.

The representative of the Eastern Algonkian languages to be examined is Passamaquoddy.

Passamaquoddy Numerals (Masthay 1977)

# Passamaquoddy	System	Micmac
1 naiget	Quinary	newt
2 nes	Quinary	ta'pu
3 nihi	Quinary	si'st
4 ného	Decimal?	ne'w
5 nanne	Decimal?	na'n
6 gamatchine	Descriptive ('across to the next finger')	asukom
7 alohegannak	Descriptive ('at the pointer')	l'uiknek
8 okmalchine	Descriptive ('finger beyond')	ukumuljin
9 askenadek	Descriptive ('one finger left')	peskunatek
10 neqdensk	Descriptive ('one set of fingers')	newtiska'q

This data shows that Passamaquoddy and Micmac numerals have similar etymologies. There are a few differences in that Micmac 2 *ta'pu* belongs to the decimal system whereas Passamaquoddy *nes* belongs to the quinary system. Both languages share a common etymology for the numbers 6, 7, 8, and 9 and possibly (but not definitely) 10. Both languages show a blending of the three numeral systems.

Some of the other numbers, apart from those of the first decade have interesting etymologies and some of these will be presented and examined etymologically in what follows. Of interest, for example, are the elements that mark the further stages of the decimal system - 100, 1000, and 1,000,000:

10 mtln

100 kask-mtln-aqn

edge - 10 - NF

kask means 'edge' as in the word *kaskipunaqtuk* 'on the bank of a river', thus indicating that 100 is the edge of ten or as will be seen from the following discussion 'the first layer of tens'.

1,000 pitui-mtln-aqn

extra layer-10-NF

pitui has the meaning 'extra layer.' Micmac *pitui-* is a reflex of PA **pi:htaw-* 'extra layer'. There are many reflexes of **pi:htaw-* in the Algonkian daughter languages: Menominee *pe:htawakotaw*, 'he hangs it double or as an extra layer'; Ojibwa *pi:tto:ya* 'have layer upon layer'; Fox *pi:htawapite:hiwi* 'it is tied with inside layer'; and Micmac *pitueka'tu* 'to line (as a coat)' and *pitueka'taqn* 'lining'. This indicates that 1,000 in Micmac means 'extra layer of tens'.

1,000,000 kji-pitui-mtln-aqn

big-extra layer-10 - NF

kji which is a reflex of PA **ke?či-* is used in many Micmac words to give them the meaning 'big' as in the following examples; *kji-mntu* 'Satan', literally 'the big devil' and *kji-pa'tlias* 'bishop', literally 'the big priest.' This Proto-Algonkian root also has cognates in other Algonkian languages as in the following examples: Cree *kisči-*, Fox *kehči-*, Menominee *ke?če-* and Ojibwa *kičči-* all meaning 'big' as in the Menominee word for 'sea, ocean' which is *ke?čekam. pitui-* 'extra layer' thus gives the meaning 'big extra layer of tens' indicating 1,000,000.

Most of the Micmac numerals can be analyzed, and have cognates in other Algonkian languages. 1 and 3 are quinary, 2, and 4 are decimal and 5 is possibly decimal or quinary. 6, 7, 8 (and possibly 9 and 10) are descriptive. 10 may also be quinary.

The whole system of numerals in Micmac is predominantly decimal-based on the formation of the multiples of ten by pitui, etc.

Micmac has a blending of the three numeral systems presented by Siebert.

REFERENCES

Hewson, John. 1987. Micmac Numerals in Betsiamites. *Algonquian and Iroquoian Linguistics*. 12.2. 18-19.

Hewson, John. in press. *A Computer Generated Dictionary of Proto-Algonkian.*

Masthay, Carl. 1977. ms

Siebert, Frank T. Jr. 1975. Ressurecting Virginia Algonquian From the Dead: The Reconstituted and Historical Phonology of Powhatan., in James M. Crawford (ed.), *Studies in Southeastern Indian Languages*, 285-453. University of Georgia Press. Athens.

PLACE NAMES IN FINLAND: SETTLEMENT HISTORY, SOCIOLINGUISTICS, AND THE FINNISH/SWEDISH LANGUAGE BOUNDARY

Sheila Embleton
York University

ABSTRACT

There is much ongoing research in Sweden and Finland concerned with uncovering settlement history in Finland through the interaction of Swedish and Finnish in place names. The Finnish case is particularly interesting, as the contact extends from prehistoric times through historic times to the present, and as the topic has become relevant to a hot sociolinguistic debate amongst scholars and the lay public alike. This paper presents a classification and exemplification of different types of loan-name found in areas of Finnish—Swedish contact, and shows how place names can provide information on age of settlements and ethnicity of various layers of settlement, using etymology and linguistic methods in conjunction with other tools (e.g., from geology). The linguistic evidence often provides fairly clear answers as to which group was original in a given area, but those answers may not be accepted by current inhabitants for political reasons. Details of the sociolinguistic debate, from both the historical and current points of view, are presented, along with an overview of current scholarly research on the topic.

1. INTRODUCTION

There is much ongoing research in Sweden and Finland concerned with uncovering settlement history in Finland through the interaction of Swedish and Finnish in place names. The basic insights from this research are also applicable to any situation of extensive historical contact between two or more populations speaking different languages. A few examples of other areas of similar contact, all the subject of extensive onomastic research, are England (Celtic, English, Norse, Norman French), the US Southwest (English, Spanish, Indian languages in for example New Mexico), Northern Quebec (English, Inuktitut), and Namibia (English, German, Afrikaans, Khoekhoen). The Finnish case is particularly interesting, as the contact extends from prehistoric times through historic times to the present, and as the topic has become relevant to a hot sociolinguistic debate amongst scholars and the lay public alike.

As is well-known, Finland is constitutionally a bilingual country, with Swedish (currently the first language of 5—6% of the total population of 5 million) enjoying the same rights as the majority language Finnish. As a result of prolonged contact between Finnish and Swedish, Finnish has a large number of loanwords from Swedish, and the Swedish dialects spoken in Finland have borrowings from Finnish. In addition there is a small minority of Saame-speakers, now restricted to the far north, and historically there has been contact with proto-Germanic and earlier forms of Baltic languages. All these contacts have had linguistic effects of smaller or larger proportions; the current paper will be restricted to Finnish—Swedish contacts (by the far the most extensive) and how these contacts are mirrored in the toponymy of Finland. But even contacts between proto-

Germanic and Finnish are reflected in a very few place-names; see for example Koivulehto (1987). The main areas of Finnish—Swedish contact (refer to the map, page 38) are in Ostrobothnia (Fi. *Pohjanmaa*, Sw. *Österbotten*, 1 on the map), the Turku Archipelago (Fi. *Turunmaan saaristo*, Sw. *Åbolands skärgård*, 2 on the map), Åland (Fi. *Ahvenanmaa*, Sw. *Åland*, 3 on the map), and the area around Helsinki (Fi. *Uusimaa*, Sw. *Nyland*, 4 on the map). The boundaries between Finnish-speaking and Swedish-speaking areas are of course fuzzy rather than exact, in that any proportion of Finnish-speakers relative to Swedish-speakers is possible in any given location, but most areas can be accurately described as either predominantly Finnish-speaking or predominantly Swedish-speaking. In addition, over the prolonged historical contact period, the boundaries (fuzzy as they may be) are unstable and have shifted, sometimes more than once even in a given area. Thus it is quite possible to find Finnish place-names where Swedish is currently spoken and vice versa. And, as will be seen below, the existence of for example a Finnish place-name is no guarantee that either the current or the original population is Finnish-speaking, but only that at some time the population has been Finnish-speaking.

2. TYPOLOGY OF LOAN-NAMES

The following typology of loan-names is based on Kiviniemi et al 1977; I present it here not for the purpose of classifying, but rather because it provides a succinct framework for giving examples of the possible types of loan-name. Other typologies have been presented by other researchers (e.g., Dorion 1972:20-21, 34-36, in the context of English/French/Amerindian in Quebec, and concerned also with questions of officialisation; Nicolaisen 1976:53-56 in the context of Scotland), but they do not differ in any important respects (Dorion is much more detailed, unnecessarily so for my purposes here). To prevent misunderstanding, I would also like to emphasize that (unlike Dorion 1972 and Nicolaisen 1976) this is a typology of the loan-names only, not of all names found in areas of contact. Thus there is no place here for a name such as that of the city of Turku (Fi. *Turku*, Sw. *Åbo*, where the two names are completely unrelated to one another). In my exemplification, I try to use essentially the same name as an example right through all 6 subtypes, to try to compare and contrast the different treatments more easily.

2.1 TYPE 1: NAMES DIRECTLY BORROWED

These are names adopted, with only minor phonetic adjustments, from one language into the other. For example for an island, an original Finnish *Mustasaari* (Fi. *musta* 'black', *saari* 'island') could become Swedish *Mussor*, or an original Swedish *Svartholmen* (Sw. *svart* 'black', *holm(en)* 'island') could become Finnish *Vartholma*.

2.2 TYPE 2: NAMES DIRECTLY BORROWED WITH EPEXEGETIC ADDITION.

Here a name is borrowed as in Type 1, but the resulting form (semantically opaque in the borrowing language) is explained as it were by the addition of another element, which will generally denote the type of place involved. The term epexegetic addition refers to this added explanatory element, which is usually a generic of some sort. For example, our original Finnish *Mustasaari* could become Swedish *Mussorholmen*, where *holmen* is the epexegetic addition. This assumes that the place is still an island when the Swedish naming takes place. But often the addition of alluvial deposits, or land upheaval, particularly common in these coastal areas of Finland, means that an original island might no longer be an island. It might for example be a cape at the time of the Swedish naming, in which case it would become Swedish *Mussorudden* (with epexegetic *udden* 'cape, point').

Epexegetic additions are of course also common outside Finland. Pellijeff (1985:458) cites a similar example from central Sweden, where Finnish died out about 200 years ago; Finnish *Kiviniemi* (Fi. *kivi* 'stone', *niemi* 'cape, promontory') has become Swedish *Kivnemsudden*. At the extreme such epexegetic additions can even repeat the whole meaning in the borrowing language. Thus, again from central Sweden, Finnish *Purohaara* (Fi. *puro* 'brook', *haara* 'forking') has become Swedish *Porroharabättjegrena* (Sw. *bättje*, equivalent to *bäck* 'stream', *gren* 'twig, branch'). One is also reminded of names in England like *Torpenhow Hill*, which have epexegetic additions like the layers of an onion: *hill* 'hill' from English, *how* 'hill' from Old Norse, *pen* 'hill' from Old Welsh, added to an original base *tor* 'hill' from Celtic. Peeling the onion gives the settlement history in neat reverse chronological order.

2.3 TYPE 3: REDUCTION PLUS EPEXEGETIC ADDITION

This is similar to Type 2, but with an elliptic or reduced form of the base. For example, our original Finnish *Mustasaari* might become Swedish *Mussholmen* (if still an island) or *Mussudden* (if a cape). It is possible that Type 2 could be an intermediate stage before Type 3 (e.g., *Mustasaari* > *Mussorholmen* > *Mussholmen*).

2.4 TYPE 4: NAMES TRANSLATED

A morpheme by morpheme translation occurs in the borrowing language. Our original Finnish *Mustasaari* would become Swedish *Svartholmen*, and an original Swedish *Svartholmen* would become *Mustasaari*; depending on historical circumstances and the availability of records, this type could of course easily go undetected.

2.5 TYPE 5: NAMES BORROWED, CONTAINING PLACE-NAMES.

These are names which contain a borrowed place-name (by any of the four types enumerated so far), plus some translated element. For example, a road leading to our island known as *Mustasaari* in Finnish would be called *Mustasaarentie* (Fi. *tie* 'road, way'). Depending on how the island is rendered in Swedish, the road in Swedish would be called *Mussorvägen*, *Mussorholmenvägen*, *Mussholmenvägen*, or *Svartholmenvägen* (Sw. *väg(en)* 'road, way'). Similarly, if a bigger and a smaller island needed to be distinguished, these might be known in Finnish as *Iso Mustasaari* (Fi. *iso* 'big, large') and *Pieni Mustasaari* (Fi. *pieni* 'little, small') and in Swedish as *Stora Mussor* (Sw. *stora* 'big, large') and *Lilla Mussor* (Sw. *liten/litet/lilla* 'little, small').

2.6 TYPE 6: RETURN LOAN-NAMES

These are names which have undergone the loaning/borrowing process twice or even more. For example, our original Finnish *Mustasaari*, first named by Finns, might have become Swedish *Mussor* (Type 1) when Swedish came to be the dominant language in the area, and later (after the original Fi. name *Mustasaari* had become extinct) Finnish *Mussori* (Type 1) when Finns again became dominant in the area. A name like *Mussori* can be easily read as exhibiting three successive waves of settlement (original Finnish, replaced by Swedish, replaced by Finnish), but of course it often is not that easy. The phonetic changes can be sufficiently great to raise many uncertainties, or different waves may unknowingly undo a previous change, effectively wiping out the evidence. For example, an original Finnish *Mustasaari* could become Swedish *Svartholmen* (Type 4) and later be taken back into Finnish as *Mustasaari* (Type 4), thus obliterating the linguistic evidence of the successive settlements. And, just to make it more interesting, this is often coupled with a comparative lack of written records. Another example shows another slight variation on

the theme. The fortress on several adjoining islands in Helsinki harbour, begun in 1747 to shore up the eastern defences of Sweden-Finland against Russia, was founded as *Sveaborg* (Sw. *Svea* 'Sweden', with an obvious nationalistic reference, *borg* 'castle, fortification'). In Finnish it was known as *Viapori* (Type 1, perfectly regular phonetic adaptation of the Swedish form). Later it became known as *Suomenlinna* (Fi. *Suomi* 'Finland', *linna* 'castle, fortification') — thus Type 4 performed on *Sveaborg*, with the variation that 'Sweden' was replaced by 'Finland', again for nationalistic reasons, but this time a different nationalism (Finnish vs. Swedish).

2.7 FREQUENCY OF EACH TYPE

Occasionally variant place forms are found. Thus for example an original Swedish *Svartbäck* has two Finnish variants: *Vartpäkki* (Type 1) and *Mustapuro* (Type 4). An original Swedish *Broby* (Sw. *bro* 'bridge', *by* 'town') has two Finnish variants: *Ruupyy* (Type 1) and *Siltakylä* (Type 4; Fi. *silta* 'bridge', *kylä* 'town'). This existence of such variants is rare however, and in no way complicates the typology, although it complicates the construction of statistics on the frequency of each type. Kiviniemi et al (1977) give the following percentages for the distribution of the six types, based on a sample of 1948 loan-names in the linguistic border areas:

TYPE	LOANS OF FINNISH ORIGIN IN SWEDISH	LOANS OF SWEDISH ORIGIN IN FINNISH
1	76.1	54.6
2	4.7	4.5
3	5.0	5.9
4	5.7	15.7
5	6.3	12.4
6	2.2	7.0

Two facts emerge clearly. First, Type 1, namely direct borrowing with phonetic adaptation, is by far the commonest type for both linguistic groups. Second, the Finnish-speaking population has translated far more names than has the Swedish-speaking population (Type 4). Kiviniemi et al suggest that this latter is because these border areas are today slowly being fennicized. I would rather suggest a simpler explanation, which will speak to both facts at once. The long-standing political and demographic situation has been, until very recent times only, that there was very little bilingualism; this would favour Types 1, 2, and 3. And what little bilingualism there was was more commonly in the natively Finnish-speaking population than in the natively Swedish-speaking population; this would cause a larger number of translations of Swedish names into Finnish by Finnish-speakers than of translations of Finnish names into Swedish by Swedish-speakers, since you have to have at least some knowledge of the other language in order to translate.

3. HISTORICAL OVERVIEW

To understand the current sociolinguistic situation, a very brief overview of Finnish history is necessary; I refer interested readers to Klinge (1987) or Häikiö (1990) for further details in English. The period of the Swedish crusades to Finland extended from about 1155 to 1500, with actual Swedish power and Christianity coming to Finland during the

1200s. Until 1809 Finland was part of the Kingdom of Sweden. The exact eastern boundary with Novogorod/Russia shifted a number of times as the result of various wars and treaties. In 1809 Finland became a Grand Duchy within the Russian Empire. The capital was shifted further east to Helsinki, from the (provincial) capital under Swedish rule, Turku, in the extreme southwest. The liberal-minded Czar Alexander I gave Finland a rather autonomous status (e.g., retention of Lutheran religion and Swedish as official language, along with more minor things like her own postal service, central bank, and customs control). The language of official business and of education during the early part of the period of Russian rule was Swedish. The first political parties were formed in the 1860s along linguistic lines: the Fennomans championing the cause of the Finnish language and the Svecomans that of Swedish. In the 1890s (the 'period of oppression'), Finland was placed under more direct Russian rule (including loss of her own postal, banking, and customs systems). This strongly stimulated Finnish nationalism, with the eventual goal of national independence, and along with it the cause of the Finnish language. Finland declared her independence in December 1917, in the wake of the Bolshevik Revolution.

With the rise of Finnish nationalism and concomitantly of the Finnish language, the language of the overwhelming majority of the population, the question of the necessity of keeping Swedish at all in any official function arose. Thus the Finnish nationalist movement gave rise to a counter-reaction, a Swedish preservation movement. And this was the impetus for some of the early place-name studies, for example those of Axel Olof Freudenthal (1836-1911) in 1866 and 1867, which included studies of all 4 areas of contact — the need to document the Swedish presence in Finland since the earliest times, as some sort of proof of a 'right to be there'. Freudenthal's toponymic studies predictably found many age-old Swedish/Norse features, such as names of Nordic gods, obsolete Swedish personal names, and other archaisms — most of which have of course been later overturned. This basic notion of naming as proof of possession is found over and over again in onomastics. It is a common concept in personal-naming as well, in for example the right of a master to name a slave, or the compulsion of a woman to adopt a husband's surname, or the renaming of immigrants with names more congenial to the dominant culture, or in some cultures in the idea that knowledge of an individual's name gives some sort of power over that individual. But naming as possession is most common in place-naming, as clearly demonstrated by the re-naming of settlements and streets that often follows in the wake of social and political revolutions (such as the former, current, and proposed renamings in many countries in Eastern Europe and in Nicaragua), and in the obsession which explorers always had with naming as part of the claiming process. To choose one such example close to home (Kerfoot 1990:333), 'since Frobisher's voyage of 1576, explorers have sailed in Canada's northern waters, naming landscape and seascape features as evidence of discovery and possession. ... Early this century, several official Canadian Government expeditions were sent north 'to show the flag', and in so doing assigned names, particularly to coastal features.' Again to choose an example closer to home, Kerfoot (1990:341) reminds us that 'Names are significant to the indigenous people [of Canada] in their land claims settlements and in establishing Nunavut ('Inuit homeland')'. It was the same basic idea of naming as possession, as indicative of early settlement and hence of a prior right of ownership, which motivated these early Swedish-speaking researchers in Finland, and still is at the root of the current debate in Finland today.

As mentioned above, there has been contact in Finland with Germanic speakers since proto-Germanic times, a fact of which early researchers were unaware. Thus a Germanic origin to a place-name does not necessarily prove Swedish origin, since proto-Germanic or some sort of proto-Scandinavian would also be possible. In addition, a careful consideration of the sound correspondences involved can sometimes separate out these various possibilities, but such care was lacking in many early studies, and, sad to say, in

many of the current contributions even by academics. The first to pay careful attention to the possibility of proto-Germanic or proto-Scandinavian origin and to careful consideration of sound correspondences was Ralf Saxén (1868-1932; main relevant publications in 1905 and 1910). Another prominent researcher at this time was Hugo Pipping (1864-1944; main publication 1918). Thorsten Evert Karsten (1870-1942; main publications 1906, 1921-3) mostly studied the place names in Ostrobothnia, and was the first to study a broader philological context, in that he himself made level measurements along the coast, allowing him to use land upheaval chronology to judge the age of a name (Huldén 1990:112). Thus a place such as *Havänden*, taken as *ändan på havet* ('sea's end'), was found to have indeed been at sea-level about 2000 years ago; Karsten took this to mean that the settlement had been founded by a Swedish/Germanic tribe 2000 years ago. He of course also acknowledged the addition of substantial numbers of Swedish immigrants during the later period of Swedish rule. It is easy for us, with a modern and linguistic perspective, to see the faulty elements in his reasoning, but at the time this was taken as proof of long-standing Swedish settlement. It should also be stated that these various scholars, particularly Freudenthal, also found numerous Finnish place names in areas inhabited by Swedish-speakers, but simply took this as evidence (supported by other historical documentation) of the continued spread of Swedish place names, particularly in eastern Uusimaa. In the early 1920s, there was an intense debate between Karsten and Oskar Fredrik Hultman (1862-1929) in the periodical *Finsk Tidskrift*; Hultman's opinion that the contemporary Swedish settlement had its roots only from colonization during the period of Swedish rule eventually prevailed.

4. CURRENT SOCIOLINGUISTIC SITUATION

This type of debate in newspapers and periodicals about the origins of place names in the areas of language contact continues to the present-day. Perhaps in some sense it becomes all the more 'desperate' as the percentage of Swedish-speakers continues to decline, and many Finnish-speakers are again seriously questioning the need for any official position for Swedish at all. The demographic decline in Swedish is quite dramatic: in 1750, 16%, in the late 1800's, 14%, in 1980, officially 6.3% (cited in Blomqvist 1990: 130). It is now virtually impossible to find any monolingual Swedish-speakers, except amongst the oldest rural inhabitants; certainly all young people are extremely competent in Finnish. 'Mixed' marriages virtually always result in a predominantly Finnish-speaking household. 'This' of course in turn decreases the need for Finnish-speakers to learn Swedish for use within Finland, and the effect 'snowballs'. Until recently, Swedish was also considered useful for contacts outside Finland but still within Scandinavia, particularly given the continuing close ties with Sweden. But even this has now changed, and more often than not conversation between a Finnish-speaker and a Swede (or even a Norwegian or a Dane) will take place in English. Coupled with the fact that the Swedish spoken in Finland is definitely non-standard from the point of view of Sweden (i.e., is quite deviant from *rikssvenska*), many are questioning the purpose of learning Swedish at all, and would like to see replacement of compulsory Swedish at school with compulsory English. Perhaps these contemporary issues contribute to the current resurgence of the place-name debate, as many members of the Swedish-speaking minority again see the need to establish their 'right' to be in Finland.

The most recent discussions have taken place in the newspapers in Ostrobothnia (e.g., *Jakobstads Tidning*, *Vasabladet*, and *Arbetarbladet*), largely occasioned by the publication of Norrman (1988), reviews thereof and rejoinders thereto. Norrman (1988) is a 536-page treatment of hydronyms in Ostrobothnia, written by a professor of English in Tampere, an expert on the style of Henry James and Samuel Butler. It seems to be built on

the following premises: 1) no name is to be ascribed to any Finnish origin; 2) connections, even of a 'look-alike' or 'echoic' nature, are to be sought even in Germany and in England (to help establish the 'Germanic' nature of the name), as well as in the lemmata in root dictionaries for Proto-Germanic and even Proto-Indo-European; and 3) the prehistoric naming pattern for hydronyms was always total pleonasm (e.g., forms such as *Lake Lake*). Obviously premise 1), the impetus from the current sociopolitical situation, is what motivates premises 2) and 3). With respect to premise 3), denying non-pleonastic naming is to deny the existence of name types like *Spruce Lake*, *Clear Lake*, *Deep River*, and so on, to give English equivalents. This would be absurd anywhere, but it is all the more absurd in Finland, where by far the dominant naming pattern in Finnish, and largely in Swedish too, is of the non-pleonastic but descriptive type. Thus, the example *Mustasaari* 'Black Island' in Finnish or *Svartholmen* 'Black Island' in Swedish (in §2 above) is totally typical of prevailing patterns in Finland. This would be immediately obvious, for Finnish at least, just from looking at any map of Finland, but the supporting statistics can be found readily in many places, the most recent of which is Kiviniemi (1990). The data and statistics specifically for hydronyms may also be found in Kiviniemi (1990:43-45). Similar statistics for Swedish in Finland can be found elsewhere, most recently in Zilliacus (1990).

Some examples of Norrman's absurdities follow, just to show the lengths to which someone will go when sufficiently motivated by these sociopolitical considerations. First, consider the lake name *Unjärv*, which fairly evidently comes from Finnish *Unijärvi* (*uni* '[dialectally] boundary marker', *järvi* 'lake'). Norrman gives this as *-järv* of unknown origin (sic — where the feature involved is a lake, in Finland, and the Finnish word for 'lake' is *järvi*!) and *un-* from *und* 'wave, water' (totally impossible by the sound laws for the relevant dialects, since *und-* should give *å(å)nd-*). A second example is *Hirvlax*, evidently from Finnish *hirvi-* 'elk' and *laksi* (modern Fi. *lahti*) 'bay, inlet'. But for Norrman, this, along with thousands of other Ostrobothnian hydronyms, is merely of unknown origin. There would be hundreds more such examples. In addition, the simple but crucial fact that most of the territory was under water at the alleged time of naming escapes Norrman. This, of course, is a non-linguistic fact, but once more illustrates the dangers of doing linguistics, even pseudo-linguistics, without a consideration of the total philological context (cf. for example Anttila 1989).

Despite all this, Norrman does not stand alone. He has received support from Karl Axel Holmberg (a docent in Uppsala) and Gösta Ågren (a Swedish-Finnish author and poet, winner of the 1989 Finlandia poetry prize). The main spokesman for the professional philologists has been Lars Huldén (1926-), himself a Swedish-speaker from Ostrobothnia who only learned Finnish while in his teens, professor emeritus of Swedish language and professor emeritus of Swedish literature at the University of Helsinki (Huldén was in the rare situation of having held both professorships), also a well-known Swedish-language poet, and president of the Svenska Litteratursällskapet i Finland ('Swedish Literature Society in Finland'). Another Finnish Swedish scholar writing under the pen-name Atlantikus (the debate has been so heated, with for example accusations of racism, that this person strongly felt the need to remain anonymous) has shown that using Norrman's 'methods' one could very easily show that Ostrobothnia was in fact colonized by the Japanese. In addition, twenty onomasticians from Sweden have published a signed rebuttal in all the major Finnish Swedish newspapers. When academics enter a field not their own, and particularly when non-academics also get involved, progress is often slowed down and the field as a whole suffers, sometimes with all work in the field being unjustly tainted. Or to quote Gelling (1990:85), who is writing in the context of similar problems in the study of British place names, 'constant difficulty is occasioned by the 'anyone can do it' attitude, which leads scholars who have no philological expertise to propound impossible etymologies.'

5. CONCLUSION

Apart from their intrinsic interest, place names can be crucial elements in revealing history, even when written records exist, but particularly in the absence of written records. Gelling (1990:85) makes this important point when she states that 'place-names constitute vital evidence for several developments in early English history, including the transition from a Celtic to a Germanic society in the fifth-eighth centuries, and the extent and character of the Scandinavian settlements of the ninth and tenth centuries.' Even in a place so well studied (some might be tempted to say exhaustively studied) and well documented as England, the contribution of the study of place names can be irreplaceable (and hence not to be left in the hands of amateurs either). This is all the more true of Finland, where so little documentation exists, and where if anything the sociolinguistic situation is (and has been) more complex. Place names can speak directly to settlement history, offering evidence that may be totally unavailable from any other source. For example, arguments of the type 'There are no Scandinavian runic inscriptions in Finland, therefore there were no Scandinavian settlements at that time, and Swedish settlement can only have been later' is based on what could be just coincidental lack of evidence, and is therefore always challengeable; place name evidence is much more secure.

The effect of such non-linguistic factors as land upheaval should not be underestimated. One obvious way in which a consideration of land upheaval would have helped would have been in totally ruling out all of Norrman's (1988) suggestions, since the land in question was underwater at the time of alleged naming. But more usefully, land upheaval data from geology can be used in the service of uncovering settlement history. In the Turku Archipelago area, land is rising at the rate of 50 centimetres per century, as the land 'snaps back', as it were, like a sponge after the pressure of the Ice Age glacial cover. In Ostrobothnia, the rate is up to 1 metre per century. Thus tracing for example the 5-metre and the 10-metre contour lines easily gives the waterlines of times past. Since the rate of upheaval is known in detail for all of these areas, it can be used for external calibration. Thus if a feature is named *-saari* in Finnish, but *-udden* in Swedish, you not only know that it was settled by Finns before Swedes (since the land only goes upwards and never downwards), but you also know the latest possible century for the Finnish settlement and the earliest possible century for the Swedish settlement. This dating by land upheaval has been used with great success by Ritva Liisa Pitkänen (1985, 1990ab) to show that Swedish settlers arrived in the Turku Archipelago only in the 12th century, that the area was already Finnish-speaking (since Finnish place names already existed and were taken over by Swedish-speakers) at that time, and that some of the Finnish-speaking settlers must have arrived as early as the 7th and 8th centuries — all in the absence of written and even archaeological records. Pitkänen's (1985) work also involves extensive on-site examination of local topography. Essentially the same ideas have been recently voiced by Gelling (1990:99), again in the British context: 'The study of place-names in relation to drift geology, together with my recent attempt to relate place-name terms closely to topography, are aspects of a new awareness that although place-name study is primarily a philological discipline the material must be studied in close conjunction with the physical environment.' In passing, I would like to note that other scholars are also beginning to consider such palaeobotanical techniques as pollen analysis (e.g., to determine whether an area has been under continual or intermittent cultivation); one such example from Scandinavia is Brink (1987).

Noting the types of name and of place named can help dismiss some previous misconceptions. For example, it was previously commonly thought that the early Finnish

presence in the Turku Archipelago and in Åland was merely seasonal, for the purposes of hunting and fishing. But work such as Pitkänen's (1990a) has shown that this is not so for the Turku Archipelago. There are names which cannot have been given by such seasonal occupants, as they denote permanent settlements and/or refer to animal husbandry (e.g., farm-names, field-names, pasture-names). Examples are names such as *Härkilot* (< **Härkäluoto*; *härkä* 'ox', *luoto* 'islet') and *Volot* (< **Vuohiluoto*; *vuohi* 'goat'). In addition, circumstantial evidence would indicate that Swedish settlers would not have taken over with mere phonetic adaptation (Type 1 above) Finnish names that were merely given by seasonal hunters or fishers, as these names, and certainly any contact between Finnish hunters/fishers and Swedish settlers, would have been ephemeral at best. As for Åland, the archaeological evidence is quite clear that by about the year 1000 the Viking era settlement had disappeared; we also know from historical sources that by the middle of the 11th century relations with Novgorod had turned sour, and Sweden was forced to strengthen her eastern defences, leading to an attempt to repopulate Åland. Hellberg (1987) shows, from the type of name given by these Swedish settlers (e.g., names in *-bolstad* 'settlement place'), that they found evidence of this earlier habitation. He is also able to show that the vacuum had drawn Finns there, in that many of the central settlements have clearly Finnish names (e.g., *Jomala*, clearly Finnish in its general shape, but also in its characteristic locative ending *-la*; *Sålis* < *Soolaaksi*, *soo* [modern Fi. *suo*] 'bog', *laaksi/laksi* [modern Fi. *lahti*] 'bay') or refer explicitly to the presence of Finns (e.g., *Finström* 'Finnish stream'; *Finby* 'Finnish village'). Further aspects of the evidence, which is too detailed to go into here, point to peaceful coexistence and some degree of bilingualism. As a final note on Åland, I should like to point out that the most recent research (Huldén 1976) indicates that the Finnish form of the name (*Ahvenanmaa*; *ahven(a)* 'perch [fish]', *maa* 'land') seems to be the original (Swedish *Å-* being the normal outcome of *ahven-* by the sound laws; Swedish *-land* simply translating Finnish *-maa*). Such evidence also points to Finnish settlement before the current Swedish settlement.

Gelling (1990:97) also notes the studies that have been done in the British context on the types of sites occupied by names (and hence settlers) of different origins: 'A major element in ... studies of Danish names in eastern England has been the plotting of English and Danish place-names on maps which show the types of soil on which the settlements stand. In some areas this reveals that villages with Scandinavian names regularly have less desirable sites than those with English names. This ... suggest[s] that actual colonisation was involved, with Danish settlers of relatively humble status breaking in land which was not being cultivated by the English.' Nicolaisen (1976:158) discusses a similar pattern with respect to Pictish settlement in Scotland, where *Pit*-names are clustered into areas of loamy soil, with good drainage and good shelter, preferably with southern exposure. Other settlers, bestowing names of other origins, had to take whatever land was left. The same types of things can be shown in Finland. For example, if names in *-järvi* 'lake' (and Swedish phonetic adaptations thereof, such as *-järv*, as seen above) are frequent but names in *-lampi* 'pond' are not very frequent, this would show that the bigger waters were named, and hence presumably occupied/settled by Finnish-speakers when Swedish-speakers arrived and occupied the more peripheral areas. It is very common in the language contact areas for there to be an interdigitation of Swedish and Finnish names (noticed already as early as by Freudenthal), which shows first a Swedish filling-up of peripheries, presumably indicating that Swedish-speakers were migrating into areas already occupied by Finnish-speakers. This may be followed then by total absorption of the Finnish by Swedish (e.g., particularly in Åland); in some areas, this may yet later be followed by fennicization in more recent times — according both to shifting currents of nationalism and of settlement history. The boundaries are both fuzzy synchronically and unstable diachronically.

'Because they have ... distribution in both space and time, i.e., geographical scatter plus linguistic stratification, place-names have come to be recognised as valuable raw material for the study of settlement history or ... of the settlement history of speakers of various languages.' These words were written in the context of Scotland's place name history (Nicolaisen 1976:34), but they could apply equally well to Finland, as indeed, no doubt, to many areas of the world where there is extensive language contact.

REFERENCES

Anttila, Raimo. 1989. 'Pattern Explanation: Survival of the Fit.' *Diachronica* 6: 1-21.

Blomqvist, Marianne. 1990. 'Swedish Family Names in Finland.' In Heikki Leskinen and Eero Kiviniemi (ed.), *Finnish Onomastics/Namenkunde in Finnland*, 130-140. Helsinki: Suomalaisen Kirjallisuuden Seura.

Brink, Stefan. 1987. 'Nordisk bebyggelseonomastik: Några allmänna och speciella synpunkter med finlandssvensk utgångspunkt.' In Lars Huldén (ed.), *Klassiska problem inom finlandssvensk ortnamnsforskning* (Studier i nordisk filologi, 67), 99-111. Helsinki: Svenska litteratursällskapet i Finland.

Dorion, Henri. 1972. 'La problématique choronymique des régions multilingues.' In Henri Dorion (ed.), *Les noms de lieux et le contact des langues/Place Names and Language Contact*, 9-41. Quebec: Les Presses de l'Université Laval.

Gelling, Margaret. 1990. 'The Historical Importance of English Place-Names.' In Jean-Claude Boulanger (ed.), *Le nom propre au carrefour des études humaines et des sciences sociales*, 85-103. Québec: Les Presses de l'Université Laval.

Hellberg, Lars. 1987. *Ortnamnen och den svenska bosättningen på Åland* (Studier i nordisk filologi, 68). Helsinki: Svenska litteratursällskapet i Finland.

Huldén, Lars. 1976. 'Namnet Åland.' In Lars Svensson et al (ed.), *Nordiska studier i filologi och lingvistik: Festschrift tillägnad Gösta Holm*, 218-224. Lund: Studentlitteratur.

Huldén, Lars. 1990. 'Place-names and onomastics in the Swedish-speaking areas of Finland.' In Heikki Leskinen and Eero Kiviniemi (ed.), *Finnish Onomastics/Namenkunde in Finnland*, 109-121. Helsinki: Suomalaisen Kirjallisuuden Seura.

Häikiö, Martti. 1990. 'Contemporary Finnish History: under Swedish and Russian Rule.' *Suomi Siltä/Finlandsnytt/Suomi Bridge* 90 (5): 30-31. [Part 1 of an 8-part series].

Kerfoot, Helen. 1990. 'Geographical Names and Sovereignty in the Canadian Arctic after 1880.' In Jean-Claude Boulanger (ed.), *Le nom propre au carrefour des études humaines et des sciences sociales*, 333-342. Québec: Les Presses de l'Université Laval.

Kiviniemi, Eero. 1990. 'Die lexicalische Grundzüge der finnischen Ortsnamen.' In Heikki Leskinen and Eero Kiviniemi (ed.), *Finnish Onomastics/Namenkunde in Finnland*, 26-46. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kiviniemi, Eero, Gunilla Harling-Kranck, Peter Slotte, and Ritva Liisa Pitkänen. 1977. 'Der Namenbestand an der finnisch-schwedischen Sprachgrenze.' *Onoma* 21: 426-439.

Klinge, Matti. 1987. *A Brief History of Finland*, fifth edition. Helsinki: Otava.

Koivulehto, Jorma. 1987. 'Namn som kan tolkas urgermanskt.' In Lars Huldén (ed.), *Klassiska problem inom finlandssvensk ortnamnsforskning* (Studier i nordisk filologi, 67), 27-42. Helsinki: Svenska litteratursällskapet i Finland.

Nicolaisen, W[ilhelm]. F. H. 1976. *Scottish Place-Names: Their Study and Significance*. London: Batsford.

Norrman, Ralf. 1988. *Några österbottniska vattennamn*. Vaasa: Scriptum.

Pellijeff, Gunnar. 1985. 'Suomalaisia paikannimiä Keski-Ruotsissa.' [Finnish place names in Central Sweden] *Virittäjä* 89: 455-470.

Pitkänen, Ritva Liisa. 1985. *Turunmaan saariston suomalainen lainanimistö* [The Finnish loan-names of the Turku Archipelago] (Suomalaisen kirjallisuuden seuran toimituksia [Proceedings of the Finnish Literature Society], 418). Helsinki & Rauma: Suomalaisen kirjallisuuden seura.

Pitkänen, Ritva Liisa. 1990a. 'The First Place-Names of a Wilderness.' In Jean-Claude Boulanger (ed.), *Le nom propre au carrefour des études humaines et des sciences sociales*, 449-458. Québec: Les Presses de l'Université Laval.

Pitkänen, Ritva Liisa. 1990b. 'Uplift as a method of dating the names of natural places.' In Heikki Leskinen and Eero Kiviniemi (ed.), *Finnish Onomastics/Namenkunde in Finnland*, 47-60. Helsinki: Suomalaisen Kirjallisuuden Seura.

Zilliacus, Kurt. 1990. 'The Place-Names in "The Skerries".' In Heikki Leskinen and Eero Kiviniemi (ed.), *Finnish Onomastics/Namenkunde in Finnland*, 122-129. Helsinki: Suomalaisen Kirjallisuuden Seura.


Barents Sea
Varangerfjord
UNION OF
SOVIET
SOCIALIST
REPUBLICS
Lapland
Monchegorsk
Apatity
Kemijärvi
Rovaniemi
Sodankylä
Pajala
Kalix
Luleå
Piteå
Skellefteå
Raahe
Oulu
Kajaani
FINLAND
Kokkola
Pietarsaari
Vaasa
Nurmes
Lieksa
Pielinen
Kuopio
Karelia
Jyväskylä
Jämsä
Tampere
Rauma
Forssa
Åland Islands
Helsinki
Gulf of Finland
Gulf of Bothnia
Lake Ladoga
Imatra
Lappeenranta
Svetogorsk
Vyborg
Kotka
Zelenogorsk
Sestroretsk
Lakhtinskiy
Leningrad
Kronshtadt
Pushkin
Petrodvorets
1
2
3
4

BETWEEN EMPHASIS AND EXAGGERATION: VERBAL EMPHASIS IN THE ENGLISH OF CAPE BRETON ISLAND

Lilian Falk
Saint Mary's University

ABSTRACT

The paper examines a variety of stylistic features used by Cape Breton speakers to emphasize a point. Among the most frequently used, we see syntactic extraposition, rhetorical openings and closings, as well as the repetition of key words and the linking of synonyms. Aside from these features, we see also the use of rhetorical superlatives and proverbial comparisons, as well as a number of rhetorical interjections. On the basis of the examples examined here, the paper tries to describe the various functions of these features in narrative segments. The analysis suggests that while some of the expressions may be specific to Cape Breton, their function in discourse follows patterns similar to those which have been observed in Standard English.

INTRODUCTION

The present paper is based on an examination of several stories and interviews in Cape Breton's Magazine, edited by Ronald Caplan. Personal stories were checked with a view to observing what features of style and vocabulary are commonly used by Cape Breton speakers to convey emphasis. Since emphasis is achieved not only by elements which are exclusive to Cape Breton English, but also by those which are shared with general English, attention is given here to both kinds.

A working approach to identifying emphasis and exaggeration will be that these are linguistic elements which belong primarily to the speaker's performance in the course of a conversation, and would be omitted in paraphrase or in re-casting into reported speech, e.g. 'Geez it is real cold out' would be paraphrased as: 'he says it is (very) cold out.' An extensive discussion of elements which are characteristic of direct speech as contrasted with indirect speech can be found in Banfield 1982.

Of these elements it can be said that, instead of conveying factual information, they convey the speaker's attitude, emotional state, and degree of involvement in what is being said. As we shall see, they also play an important part in establishing the speaker's authority as a reliable source of information.

1. MEANS OF EMPHASIS

1.1. EMPHASIS BY SYNTACTIC EXTRAPOSITION AND WORD-ORDER

In addition to words which would be omitted in paraphrase, there is also an emphasis achieved by sentence structure or special word order, which would be recast in paraphrase. 'With my cross-bow I shot the Albatross' is likely to be paraphrased as 'he shot the Albatross with his cross-bow'. Here, words are not being omitted but order of elements in the sentence is recast. Therefore, syntactic rearrangement or extraposition of sentence elements will be regarded as a means of emphasis.

Since instances of syntactic emphasis may come in the same utterance with exaggeration such as high numbers or expressions of intensity, they cannot be always treated separately, but we will look at some examples.

In creating emphasis by the arrangement of syntactic elements, the speaker achieves the prominence of one component.

(1) This young fellow and I, we borrowed a lantern from the fellow...

(Caplan 1984b (henceforth 36) :13)

This sequence puts the subject, *this young fellow and I*, into prominence. In the following sequence a speaker wishes to emphasize the point that his mother played a greater part in getting him to do responsible work when he was still a young boy than his father did.

(2) Don't forget, our mother had a lot to do with this. *It was her* that got us to do the work.

...And *it was* really her that got us to do the work.

...And *it was her* that really trained us... But *it was* really her that got us to do the work.

...And *it was her* that would get us up in the morning. 36:10

(3) It used to be the mother that used to say, 'Now, your father needs you tomorrow morning...' 36:10

In examples 2 - 3 the emphasis is achieved by the device also common in general English: placing the subject in a main clause, with the significant verb in a subordinate clause.

In addition, *really* is also used to stress the significance of the verb, and finally the entire effect is still strengthened by having the pattern repeated five times in a short connected sequence in a conversation.

The use of the pronoun form *her* as emphatic subject is also consistent with the emphatic paradigm *it is me/him/her/us/them.*

(4) Very *seldom* would I kill a calf. 36:11

Here the adverb is put in the initial position for emphasis. The adverb *seldom* is further intensified by *very*, and *would I kill* also shows emphatic order.

(5) I had gone on the ice, and *down* she went. 36:14

The position of *down* before, rather than after the verb, creates the emphasis.

Thus it can be said that stylistic features employed to create emphasis are used not in isolation, but in combination with a variety of such features. Examples of these will be considered further on.

1.2. RHETORICAL OPENINGS AND CLOSINGS

If by rhetorical openings we mean those discourse markers which serve to signal the fact that the speaker is beginning, resuming, or changing a point of view, or answering a question, then, for the most part, we see signals shared with general English, such as *but, and, of course, well, see, as a matter of fact, so, anyway,* and *so anyway*, and the like. In medial positions and in closings we also note common expressions like *you know, my God, geez, oh geez*, and others more specific to the region, like *Holy God almighty.*

(6) *Yes indeed* I was in the dory when they [the swordfish] came through.
(Caplan 1973 (henceforth Coll. Ed.) 4:6)

(7) *You see* if you died there you had to wait until summertime to be buried.
Coll. Ed. 5:15

(8) *Well, look*, I had that card in my clothes box for I don't know how many years.
Coll. Ed. 3:A1

(9) *Well, I'll bet* it's been 45 years since I didn't make that. Coll. Ed. 4:17

10) *And* I remember the night he died. Coll. Ed. 5:0

CLOSINGS:

(11) Violin places are using it yet *anyways.* Coll. Ed. 4:17

(12) ...somebody would stick some — there would be fish there *all right.* Coll. Ed. 4:3

OTHER CLINCHERS:

(13) *You know*, I think I'm the happiest woman on earth. Coll. Ed. 5:25

The openings and closings serve as links between parts of the discourse, and help to establish the speaker's control of the narrative, as well as his/her awareness of the control and the attending responsibility. *You see*, above, introduces an explanation to the hearer, not a part of the story as such.

1.3. DISJUNCTS

But the control and responsibility also take other forms. A speaker can testify to his/her honesty, accuracy, frankness and similar aspects of involvement with his/her own statement. Such testifying can take very strong forms, and can easily exceed the measure of whatever the situation vouchsafes objectively, if objective measurement could be applied. The testifying can also take the form of a caution, indicating that the speaker's intent is to be as accurate as possible, but he/she fears that his/her memory or his/her powers of observation are not fully reliable.

While Quirk et al. (1985) list adverbs such as *frankly, honestly, simply* as common disjuncts, our texts show the use of parenthetic clauses and phrases having a function similar to such adverbs. Stubbs (1983: 70) speaks of such adverbs as making 'metareferences to the discourse itself'.

CLAUSES WITH FIRST PERSON PRONOUN AS SUBJECT:

(14) Snowing? You couldn't see your hand before you. *I'm telling.* 36:70

I'm telling carries some of the force of stressing the severity of the snow storm. It can be used only in accounts of first person observation, creating a link between the fact observed, and the person of the speaker.

A link also can be created between the fact observed, the speaker, and the listener:

(15) Well, people ask, how'd you live? *I'll tell you how we lived.* A hell of a lot better than people live today. 36:68

These assertions of the speaker's reliability also heighten the significance of the speech act as an important event in itself, as if the event of the past was receiving a renewed significance through being spoken about:

(16) *I am safe in saying* I worked there for three years for Robin Jones ... 5 cents an hour. 36:68

In addition to clauses, speaker's assertions can appear as phrases:

(17) *Not blowing or bragging about it,* I was reared ... the toughest kind of a way. 36:72

or:

(18) *...but as far as my part,* there's not a quarter of the company. 36:72

A special way of involving the hearer in the assertion of the speaker's reliability is the use of impersonal *you.* Sufficiently ambiguous as to whether it refers to the speaker or to the hearer, it suggests a reference to both:

(19) ...[we] saw that boat coming around Middle Head, and *you'd swear* she was one of those great big slabs of drift ice... 36:72

The difference between disjuncts testifying to the speaker's reliability and seemingly objective claims of the truth of a statement may sometimes not be easy to set apart.

(20) That's what's causing cancer today. Because we're not getting the right kind of food — *and that's the whole truth of it..* 36:68

Here the form is that of an objective claim, but the claim is in reality a personal opinion. Presenting personal opinions as general truth, or endorsing a general claim by personal testimony, are both frequent in personal narratives, further suggesting the extent to which speakers perceive their responsibility, not only in personal, but also in general matters.

Certain assertions may have the force of a denial, as in:

(21) *Really truly,* I wasn't frightened, but I thought we were going to be drowned. 36:71

Here the positive assertions *really truly* is used to confirm the truth of a statement, which in itself is a denial. Statements beginning with never, and other negations are treated below.

1.4. EMPHATIC DENIALS

Negative statements are no less subject to emphasis and exaggeration than are assertions. For a speaker, it is as important to establish his/her trustworthiness in matters of denial, as it is in matters of assertion. It is, however, a little more difficult, as the resources are more limited. Multiple negatives, for instance, are discouraged. The expression 'he never yet no vileinye ne sayde ... to no maner wight' is fondly recited by scholars, but has little currency in everyday speech. Denials in Cape Breton speech depend mostly on repetitions of certain elements. Below are some examples:

(22) You *never* were cold, *never.* 36:69

(23) *Never* in my life, *never.* 36:69

(24) You *never* were cold that way, no way in the world. 36:71

(25) I *never* stopped fishing, no sir. 36:70

(26) (Must have been cold.) Yeah, but you'd *never* mind out there. Wouldn't mind that at all. 36:69

The element of exaggeration in the above examples can be appreciated in the context, for in fact the narrative stresses that the weather was very cold, but it was endured without complaint, and the denial *never in my life, never*, is simply a prelude to a dramatic account of the single occurrence of the incident in question. The speaker was caught on a fishing hook once, and remembers the painful incident vividly.

No way, or the even more emphatic *no way in the world* are used when the denial is one of manner (as distinct from denial of event in actual time):

(27) And then you couldn't see 50 feet — *no way in the world.* 36:71

(28) You had your winter clothes on you — and once you put your oilclothes on, brother, there is *no way* water can go through them. 36:69

Absence of a certain condition may also find a strong proverbial expression, as in:

(29) [it was] a calm day, *not a hair of wind.* Coll. Ed. 4:24

The strength of these negative assertions is the more striking, as they are not brought out by any personal provocation, but serve to bolster the speaker's position of trust and authority.

1.5. INTERJECTIONS

The words *boy, brother, son*, are often used as emphatic interjections. *My boy* and *my son* also occur, as does *my dear, my dear man*, and *my God almighty*. Of these, *my dear* is not gender-specific, and while *boy, son*, are masculine, their use is not limited to speech directed to male persons. General use also employs these words, but in Cape Breton speech they occur more often. Their function is also different from general use where *brother* is sometimes used as an exclamation of annoyance, perhaps a polite version of *bother.*

It appears that *boy*, *brother*, *son* are rarely used at the opening of a conversation segment. Rather, they tend to come medially or as the final element, after some tension has been built up, and *boy* etc. helps to achieve the climax.

(30) The least thing will take them up, but when they come up, *boy*, watch out. 36:64

(31) And, *boy*, the hook got me there. 36:69

(32) (How did you avoid the hooks while hauling?) You got used to it just like everything else, *boy*. 36:69

(33) And that Huey MacKenzie we were talking about, that Huey MacKenzie who had the Gaelic and everything, had a book, *boy*, stories — this knock came one night and the door opened and this hand came on the wall and Huey was sitting somewhere and he had a pillow and he just aimed on the hand — and the hand, *boy*, threw the pillow back in his face. That was going on, *boy*, and she was to priest and she was everywhere — no use. Coll. Ed. 6:0

Brother is similar in use to *boy*; in the example below it seems to intensify the negative *no way*:

(34) You had your winter clothes on you — and once you put your oilclothes on, *brother*, there is no way water can go through them. 36:69

My boy, *my son*, *my dear* act as a somewhat weaker emphatic element than *boy*.

(35) You'd never stop, *my boy*, - - 10 hours. 36:68

(36) Then the last going off, well, there's fish, *my son*, it takes two with a gaff to get in the dory. 36:70

Dear man occurs along with *man dear*, and *dear*.

(37) Oh, *dear man*, that roller's as good as two men. 36:65

(38) You had a wonderful sou'wester buttoned underneath your chin, and *man dear*! (You didn't haul barehanded?) You couldn't, *dear*. 36:69

Dear man, *man dear*, and even *my dear* resemble but are not identical with the use in general English, where *my dear* is more likely to be used to a woman listener as a form of rhetorical address than to a man. In Cape Breton speech it is an emphatic element, which serves to highlight the constituents nearest to it.

(39) I've seen us out there, *my dear*, in the Beatrice, after my father sold the Whitty boat. 36:70

(40) We fished till the drift stopped us — all the young people. But, *my dear*, there's an awful difference. 36:72

My God almighty provides a strong element:

(41) An awful lot of work to them, but they were great boats, *my God almighty*. 36:67

Jeez is perhaps a little weaker:

(42) If I've got one anyways big or something, *jeez*, if I tried to hold it it'd tear me to pieces. Coll. Ed. 6:5

(43) In an hour or so a fellow says go on up now and see what she's like. *Jeez*, I went in — she was lying on the lounge with an apron over her head. Coll. Ed. 5:15

Syntactically, these expressions come not only between such separate elements as clauses and phrases, but also in the middle of a syntactic construction, e.g. between subject and verb, as in 33 'and the hand, boy, threw the pillow back in his face'.

2. INTENSIFIERS

In drawing a distinction between emphasizers and intensifiers, Quirk et al. (1985) reserve the term **intensifier** for those elements which add force to gradable constituents. Thus *very* in *very good* is seen as intensifier.

In our examples, we see certain intensifiers with adjectives of size and quality, including adjectives which are themselves a part of an adverbial phrase. In our texts, we see that adjectives can be intensified by repetition of the adjective itself, or by the use of a synonym. *Big* is frequently so used; *little* is rare. The following example shows repetition of *big*.

(44) [They] even went onto the Grand Banks swordfishing. They took *big*, *big* fish. Coll. Ed. 4:4

Below are several examples with *big* being intensified by *great*:

(45) We knew they [the fish] were going to leave, because of the *great big* fish we got. 36:70

(46) ...and you'd swear she [the boat] was one of those *great big* slabs of drift ice. 36:72

(47) *Great big* pair of rubber boots on us and our oilclothes — walked over 8 miles.36:71

(48) But hauling gear, you had to haul with a *great big* pair of woollen mitts on. 36:69

The following example shows the use of *great big*, as well as *big* standing alone, in the same sentence:

(49) And there was a *great big* square-rigger going into Englishtown. And there was a *big* sea on, a *big* roll, you know. 36:69

As in general usage, *big* can be intensified by *very*:

(50) And there's *very big* sores on the side of them — and that's the sign of the last school of fish. 36:70

The last example shows *very* as intensifier, before a constituent which differs somewhat from the ones mentioned earlier. The constituents which were modified with *big big*, or *great big* can usually be seen to be concrete and, in a sense, stable in size, like a

fish, a square-rigger, a pair of rubber boots and even a slab of ice, whereas a *big sea,* and *big sores* are not stable in size. *Tremendous* sea, and *quite* a sea, show other expressions used to underscore the size and force of the gale. An abstract term, like difference, is modified with *great* only:

(51) *great* difference. Coll. Ed. 6:21
(awful difference)

Although there is also a distinction between *big* and *great* in general English, with *big* being favoured as qualifier of concrete nouns and *great* with abstract, (e.g. big house, great improvement) the distinction is not strict, and *big difference* is possible. What we see here is, within our corpus, the absence of *great big* as modifier for an abstract noun.

The form of a superlative adjective can have the function of an intensifier, as follows:

(52) the *healthiest* life on earth 36:70

where *on earth* forms a part of the intensifier.

(53) the *toughest* kind of a way 36:72

(54) in the *hardest* kind of a way 36:69

where the entire expression functions adverbially.

(55) The *very best* stuff 36:68

where *very* functions as intensifier to *best,* in itself an expression of highest quality.

As much as a superlative can be established by being the best or greatest anywhere, so it can be established as the best or greatest ever, or in the speaker's experience:

(56) That was about the *worst* experience that ever I had on the water. 36:72

In Cape Breton, as in general English, also the comparative sometimes requires a strong reinforcement, as in:

(57) *a hell of a lot* better than... 36:68

Example 56, however, is not a pure intensifier, as it is also modified by an approximator, *about.* Such seeming inconsistency is considered separately.

2.1. APPROXIMATORS AND COMPROMISERS

In discussing the function of adverbs as intensifiers, Quirk et al. (1985) open also a subclass of **downtoners**. Downtoners may seem to be too far removed from intensifiers to be included in this part of the discussion, yet it is possible to see them as points on a continuum. All the same, how can an approximator or even diminisher (e.g. *partly*) function in conveying emphasis? The answer seems to lie again in the connection between emphasis and the need of the speaker to establish himself/herself as reliable narrator. Thus hesitations, concessions, expressions of approximation and even inexactitude give an aura of genuine effort to be as accurate as possible. Some forms of approximation also involve the hearer, when the hearer is called upon, as it were, as participant in evaluating the degree

of accuracy. The parenthetic *like*, *kind of*, and *about* occur in Cape Breton as they do also in general English. Both *kind of* and *like* can occur in one sequence:

(58) And on the roll of the sea, *like*, if you are rowing you've got to kind of back water — don't let your dory go ahead too fast. 36:69

(59) ...my father would haul *like* three or four lines to give the other fellow a chance to warm up, you know. 36:69

Here *like* expresses the approximation, and *you know* involves the hearer, as does also the following statement of concession:

(60) sometimes when it was cold — *what you call* cold — 36:69

An approximator may be also used with a very strong intensifier:

(61) That was *about* the *worst* experience that ever I had on the water. 36:72

or with an estimator of measure:

(62) *about* three foot six or something like that Coll Ed. 6:16

In (61) the expression of approximation strengthens the intensifier contained in the superlative by showing that the speaker is aware of his own hesitation, but goes on to express the superlative in spite of it.

As high as is used when numbers are approximate, but at the same time convey the notion that the number is relatively high, thus:

(63) In the evening we'd have *as high as* 16 to supper. Coll. Ed. 5:19

(64) And there'd be *as high as* 70 men working there. 36:68

Giving the exact time, size, or location pertaining to an event is also a part of stressing the function of the speaker. This often comes in the form of *right* as in:

(65) Yes, indeed, I was in the dory when they [the swordfish] came through the dory *right* to their eyes. Coll. Ed.4:6

or

(66) The only time that ever I got a hook stuck into me was in that finger *right there*.36:69

(67) And the boats came down from everywhere. *Right* from Digby, *right around* down Nova Scotia and all down the shore. Coll. Ed. 4:2

(68) It will rot *right* in no time. Coll. Ed. 6:8

In this adverbial use, *right* conveys precision, from the point of view of the speaker, which may not communicate much to the hearer, except in (68) where the speaker points to the precise spot.

There is also another use of *right* as intensifier as in *right good* or *right hot* which is similar in function to *very* in general English, but it differs in that *very* can be repeated, as in *very very good*, and *right* cannot; *right* also belongs rather in first person narration, while *very* has much wider use.

2.2. AMOUNTS AND QUANTITIES

Exaggerating numbers and amounts is a common feature of everyday conversation. 'Thousands of ants', 'millions of mosquitoes', will liberally embellish an account of a less than perfect picnic. And yet even this feature of conversation, so common that it can usually pass unnoticed, has an important function. It is much more likely to occur in first-person narratives, than in second-hand narratives, (e.g. we didn't stay long, because of those millions of mosquitoes vs. they didn't stay long because the mosquitoes bothered them.) In other words, the amounts and quantities have a share in establishing the narrator's fortitude or endurance, or some other outstanding quality. They also, by the same token, establish the narrator's trustworthiness, for participants in events acquire a special right to exaggerate.

In Cape Breton, quantities can be overstated:

(69) (Did you ever have it happen?)

Yes, *thousands* of times. 36:64

[That is, the painful experience of getting squid 'ink' in the eyes]

(70) Used to go to the Bird Islands. Used to go up there to get bait, to bait our trawl gear. We'd stay there in the night. (What kind of bait?) Squid. *Thousands* and *thousands* of them. 36:64

(71) You had your vegetables, everything. Lots of meat, *piles* of meat, *piles* of potatoes, turnips — all kinds of stuff to eat. 36:68

Here *piles* means, evidently, 'large quantities' while in general English it is normally used literally for items that can be stacked or piled or, metaphorically, for money.

An expression of somewhat unusual nature was:

(72) I wish we had a *scad* of them right now. 35:10

Since an overstatement has in it an element of subjective evaluation of an objective fact, it is unexpected to see it in a putative situation where there is no objective fact which may receive such an evaluation.

3. CONCLUSION

When we look at various ways in which speakers introduce emphasis into their discourse, or intensify certain elements in a given region or speech community, we see a mirror of what obtains in general English. The desire to establish one's position as trustworthy and reliable reporter of events, and perhaps the pleasure of occupying such a position is universal. The choice of words and expressions used to this effect may, on the other hand, be different in different regions. The speakers of Cape Breton have, in addition

to the shared resources of linguistic emphasis, a number of expressions which are regional. This paper has looked at a sampling of such expressions.

REFERENCES

Banfield, Ann. 1982. *Unspeakable Sentences. Narration and Representation in the Language of Fiction.* Boston: Routledge and Kegan Paul.

Caplan, Ronald, Ed. 1973. *A Collector's Edition of Cape Breton's Magazine.* Wreck Cove, Cape Breton, Nova Scotia.

Caplan, Ronald (ed.) 1984 a. Cape Breton's Magazine, No. 35.

Caplan, Ronald (ed.) 1984 b. Cape Breton's Magazine, No. 36.

Quirk, R., S. Greenbaum, G. Leech, & J. Svartvik. 1985. *A Comprehensive Grammar of the English Language.* London: Longman.

Stubbs, Michael. 1983. *Discourse Analysis. The Sociolinguistic Analysis of Natural Language.* Oxford: Basil Blackwell.

GÉOLINGUISTIQUE SOCIALE D'UNE VILLE ET RECHERCHES COMPARATIVES

Nathalia Golubéva-Monatkina
Institut des Relations Internationales
Moscou

RÉSUMÉ

La géolinguistique sociale d'une ville a le but d'analyser la variation sociale du langage dans l'espace d'une ville. J'ai choisi le terme géolinguistique sociale *mais il serait possible de désigner cette branche de la dialectologie sociale par les mots* géographie sociolinguistique *ou* géosociolinguistique *et même par les mots tels que* géolinguistique sociale et ethnique. *Cette large variabilité du terme est conditionnée par le fait que la variation du langage dans l'espace d'une ville dépend des facteurs sociolinguistiques, linguistiques, ethniques, géographiques sans parler des facteurs historiques, économiques etc. Il y a aussi le caractère de la structure urbaine, la croissance territoriale de l'agglomération urbaine, la part de la migration intérieure et de l'immigration étrangère dans la formation de la population urbaine, une politique de langue déterminée par la nature socio-économique de la société qui doivent être pris en considération dans l'étude linguistique de la ville.*

Le paysage sociolinguistique d'une ville correspond *grosso modo* à la manière qu'ont ses habitants de s'étendre sur le territoire de leur ville, mais il nous semble évident qu'il n'existe pas d'isomorphisme complet entre la manière des habitants de répartir socialement et ethniquement le territoire de la ville, d'un côté, et la géolinguistique sociale de cette ville, de l'autre côté.

Pour peindre le paysage sociolinguistique de telle ou telle ville il est nécessaire de connaître sa morphologie sociale dans une présentation cartographique et le système des variétés littéraires (codifiées et non-codifiées) et non-littéraires de la langue (des langues) parlée(s) dans cette ville.

Il est naturel que la morphologie sociale de différentes villes soit différente, mais pour la géographie sociolinguistique semble importer un seul trait de différence – il importe de savoir si la morphologie sociale cartographiée est mosaïque ou si elle a un caractère 'de secteurs'. En regardant les cartes sociales de deux villes françaises (Reims, dessin 1, et Amiens, dessin 2) et d'une ville de l'URSS (Kazan, dessin 3) on voit que les villes françaises ont des quartiers socialement compacts, c'est leur grande ressemblance, et leur différence est minime du point de vue qui nous intéresse. Quand à Kazan, on voit que ses cellules territoriales ne forment pas de quartiers compacts socialement, mais la morphologie sociale de cette ville est mosaïque, ce qui la diffère de Reims et d'Amiens.

Pour illustrer l'importance du système des variétés sociolinguistiques dans le paysage linguistique d'une ville il est utile de voir de plus près ceux qui fonctionnent à Moscou et à Paris. Ces deux villes possèdent une grande diversité langagière qui se révèle dans la coexistence de variétés littéraires et non-littéraires de la langue nationale. A Moscou il existe deux variétés littéraires (ce sont le russe littéraire codifié, qui est langue de

littérature, d'ouvrages scientifiques, de conférences, de conversation dans toutes les situations officielles, et le russe familier utilisé dans une ambiance de famille, d'amis, dans toutes les circonstances où la communication est sans aucune contrainte de circonstances 'officielles') et une variété non-littéraire (le russe populaire qui ne forme pas non plus de variété unie). Quant à Paris, il nous semble possible d'y constater la présence de trois variétés littéraires. Ce sont: le français littéraire, langue de littérature, soigné, soutenu à un certain niveau, existant sous une forme écrite [thèses, articles...] aussi bien qu'orale [exposés, rapports, conférences...]; le français courant qui est utilisé, sous formes parlée et écrite, quotidiennement dans la communication plus ou moins officielle; le français familier); et une variété non-littéraire (le français populaire). Les domaines d'emploi du russe littéraire codifié correspondent à ceux du français littéraire et du français courant, les statuts sociolinguistiques du russe familier et du français familier semblent coïncider, de même que ceux du russe populaire et du français populaire.

Ayant la morphologie sociale cartographiée d'une ville et le système des variétés sociolinguistiques parlées dans cette ville le chercheur peut peindre le paysage linguistique urbaine avec plus ou moins de précision. Tout en prenant en considération l'absence de l'isomorphisme absolu des tableaux socioethnique et linguistique, ce chercheur semble pouvoir affirmer que dans les villes du type de Reims ou d'Amiens le paysage linguistique a un caractère 'de secteurs' tandis que celui de Kazan est mosaïque.

Au commencement des années 80 il était décrit la morphologie sociale de Paris (Golubéva-Monatkina et Roukavichnikov 1986), l'étude basée sur les données du prix de l'achat et de la location des logements neufs et anciens (Frémy et Frémy 1981). Ces données étainent traitées à l'aide de l'index de Lieberson (v. Robinson 1980). La carte sociale de Paris, qui est le résultat des recherches entreprises, permet de 'lier à l'endroit' (avec plus ou moins de précision) les variétés sociolinguistiques parlées à Paris d'il y a dix ans: dans les 'grands' quartiers c'est plutôt le français littéraire codifié et non-codifié qui est parlé; dans les quartiers 'populaires' ce français est moins utilisé, c'est plutôt le français populaire qui est parlé (il ne faut pas oublier que ce français n'est pas du tout une variété unie); il y a encore les quartiers à domination des ouvriers immigrés qui ont leurs propres problèmes langagiers.

La carte sociale de Paris (dessin 4) a un caractère bien prononcé 'de secteurs' et le paysage sociolinguistique de cette ville a aussi ce caractère. Bien qu'elle semble respirer de 'sociologie vulgaire', cette constatation n'en reflète pas moins ce qui existe en réalité. Ce caractère 'de secteurs' du tableau linguistique de Paris se fait plus évident si on le compare à celui de Moscou qui est plutôt mosaïque, en tout cas beaucoup plus mosaïque et moins 'de secteurs' que celui de Paris (à notre connaissance, la carte sociale de Moscou n'existe pas encore).

L'optique de nos recherches donne donc encore une possibilité d'études comparatives: elle permet de comparer les paysages sociolinguistiques de deux ou plusieurs villes et de se rendre compte qu'ils ne sont pas toujours semblables. Pour ce qui est de Paris et de Moscou, il est assez curieux de noter que cette différence de 'paysages' ne surgit qu'après 1917 quand Moscou est devenu 'mosaïque'. Au XIX^{e} siècle les cartes de Moscou et plusieurs ouvrages de référence attestent l'existence des quartiers compacts socialement et ethniquement. La ressemblance du Paris actuel et de Moscou avant 1917 allait beaucoup plus loin – jusqu'à la même quantité et la même disposition des 'parties' de Moscou et des arrondissements de Paris.

Et encore une possibilité d'études contrastives: on peut comparer deux paysages sociolinguistiques d'une même ville à deux moments différents. Après 1917 Moscou a très

vite perdu son paysage 'de secteurs' pour avoir celui où les cellules sociales et ethniques sont dispersées d'une manière plus ou moins égale sur tout le territoire de la ville.

RÉFÉRENCES

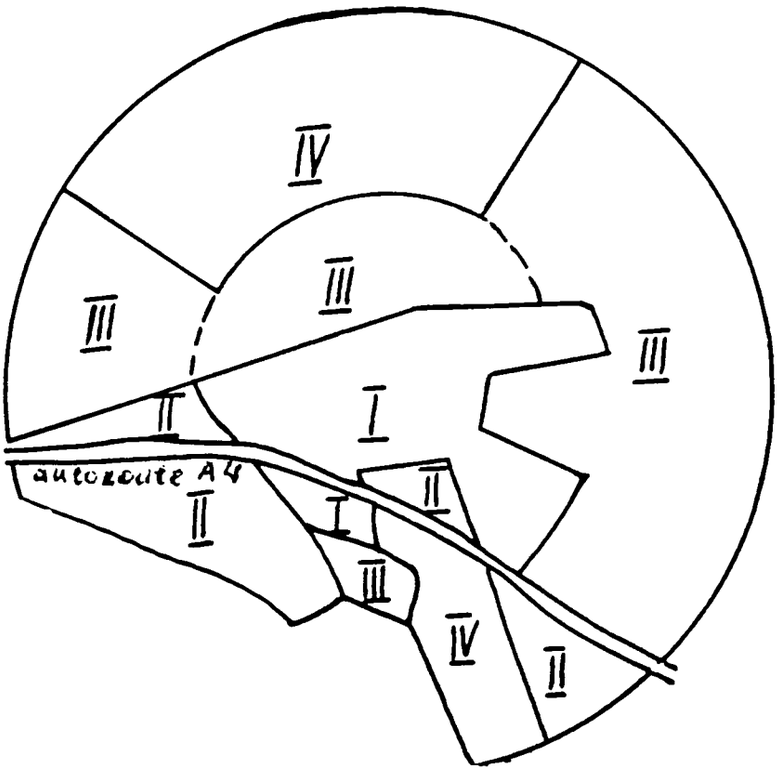
Debonneuil, X. et M. Gollac. 1978. 'Structure sociale des villes.' *Economie et statistique* 98.

Frémy, D. et M. Frémy. 1981. *QUID 1982*. Paris: Robert Laffont.

Golubéva-Monatkina, N. et V. Roukavichnikov. 1986. 'Quelques aspects de l'image sociale de la capitale française.' *Sociologitcheskjé isslédovanja (Recherches sociologiques)* 4: 160-167 (en russe).

Robinson, V. 1980. 'Lieberson's isolation index: a case study evolution.' *Area* 4:307-312.

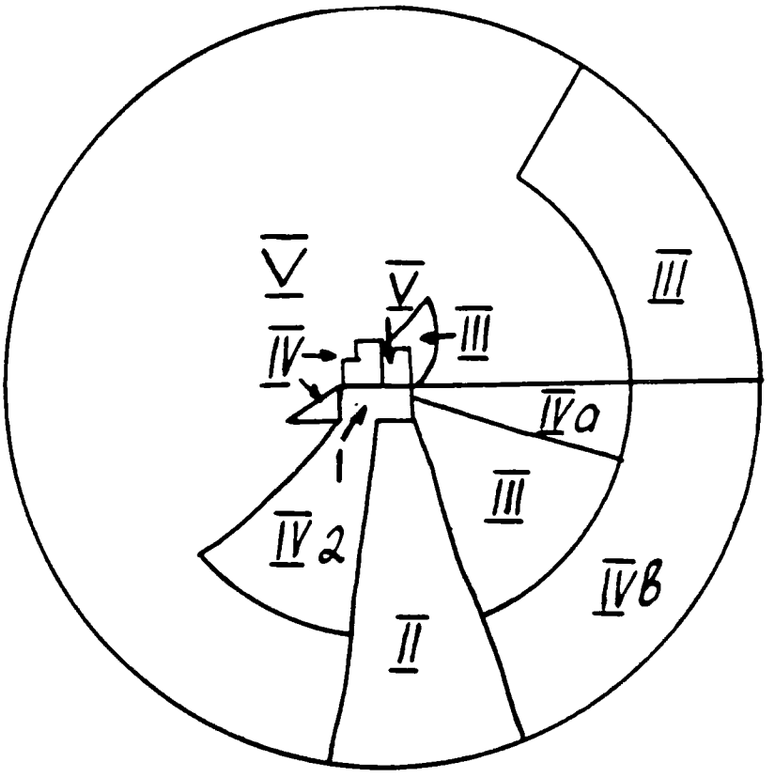
Roukavichinikov, V. 1980. *Nasselenie goroda (Socialnij sostav, rasselenjé, otchenka gorodskoj sredi)* = [*Population urbaine (Structure sociale, répartition territoriale, évaluation du milieu urbain.*] Moscou: Editions de statistiques. (en russe)



Dessin 1. CARTE SOCIALE DE REIMS
(Debonneuil et Golac 1978)

I - centre "bourgeois" et de commerce
II - quartiers des milieux "moyens" et "hauts"
III - quartiers ouvriers traditionnels
IV - quartiers ouvriers récents

CARTE SOCIALE DE REIMS



Dessin 2. CARTE SOCIALE D'AMIENS
(Debonneuil et Golac 1978)

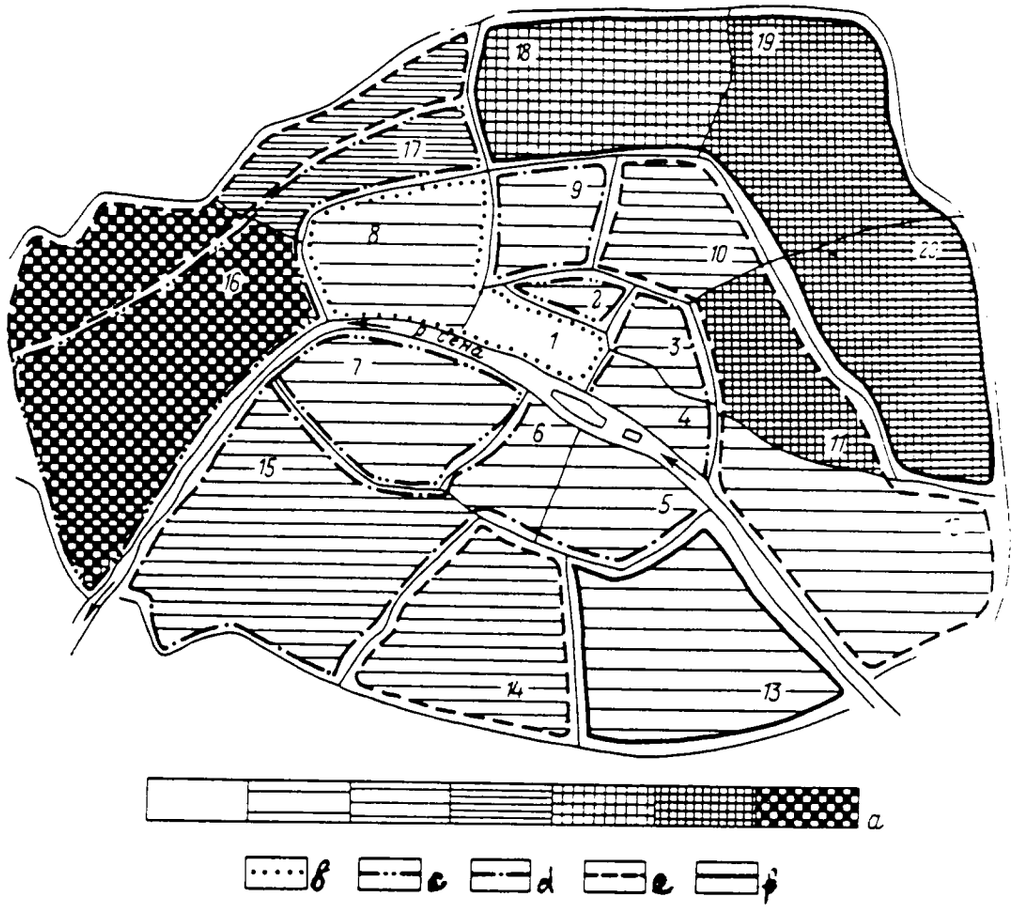
I - centre de commerce, "bourgeois"
II - quartiers bourgeois résidentielles
III - quartiers des milieux "moyens"
IVa - quartiers populaires traditionnels et quartiers des "milieux moyens"
IVb - faubourgs à dominante d'ouvriers qualifiés
V - faubourg à dominante d'ouvriers spécialisés et manoeuvres

CARTE SOCIALE D'AMIENS

Dessin 3. CARTE SOCIALE DE KAZAN
(Roukavichnikov 1980)

I - bonnes conditions de logement; quartiers à dominante d'ouvriers
II - bonnes conditions de logement; quartiers sans dominante d'une "couche" sociale
III - bonnes conditions de logement; quartiers à dominante d'employés
IV - bonnes conditions de logement; certaine dominante d'ouvriers
V - conditions de logement "suffisantes"; quartiers sans dominante d'une couche sociale
VI - conditions de logement "suffisantes"; certaine dominante d'employés
VII - conditions de logement mauvaises; quartiers ouvriers
VIII - conditions de logement mauvaises; quartiers sans dominante d'une couche sociale
IX - conditions de logement mauvaises; quartiers à dominante d'employés
X - conditions de logement mauvaises; certaine dominante d'employés

CARTE SOCIALE DE KAZAN



Dessin 4. CARTE SOCIALE DE PARIS

a - intensité des significations de l'index de Lieberson pour les immigrés habitant Paris;
b - centre bourgeois;
c - quartiers bourgeois;
d - quartiers des milieux bourgeois et "moyens";
e - quartiers des milieux moyens et des ouvriers;
f - quartiers ouvriers
(b-f marquent les frontières des zones; les numéros désignent les arrondissements.)

CARTE SOCIALE DE PARIS

QUESTIONS WITH PREVERBAL SUBJECT IN DOMINICAN SPANISH: MORPHOLOGICAL VARIATION AND SYNTACTIC RESTRUCTURING

David Heap
University of Toronto

ABSTRACT

Many Caribbean Spanish dialects tolerate preposed subject pronouns in questions where standard varieties of Spanish require subject inversion. Thus questions of the type ¿Qué tú quieres? *are ungrammatical in Standard Spanish, but completely grammatical, indeed normal, in Cuban, Puertorican and Dominican. This phenomenon can be explained as the result of subject pronouns cliticisation. New data from a recent study of Dominican Republic Spanish begin to address issues such as exactly what elements can be cliticised, and under what conditions. They also suggest a hypothesis about the interaction between morphological variation and syntactic restructuring.*

1. INTRODUCTION

1.1. SOME FACTS AND A HYPOTHESIS

A characteristic feature of Caribbean Spanish dialects (hereafter simply CS) which has been pointed out by many authors, beginning with Kany (1945), is the occurrence of *wh*-interrogative sentences with subject pronouns placed after the initial *wh* phrase and before the verb. Questions like those in (1) are attested in Cuban, Puerto Rican and Dominican dialects, and are ungrammatical in other Spanish dialects, be they European or Latin American. The questions in (1) contrast with their Standard Spanish equivalents in (2), which have either null subjects or postverbal subject pronouns.

(1) a. ¿Qué ustedes quieren? (Harris & Vincent, 1988:116)
'what you (plural) want?'

b. ¿Dónde tú vives? (Burunat, 1987:216)
'where you (singular) live?'

c. ¿Qué tú dices? (Henríquez Ureña 1975:232)
'what you (singular) say?'

(2) a. ¿Qué quieren (ustedes)?
'what want (you-plural)?'

b. ¿Dónde vives (tú)?
'where live (you)?'

c. ¿Qué dices (tú)?

'what say (you)?'

For the purposes of this paper, 'Standard Spanish' (hereafter, SS) will simply refer to all those dialects in which questions like those in (1) are ungrammatical. In *wh* questions, SS requires that a subject pronoun, if present, be placed after its verb[1]. It is immaterial for present purposes just how this phenomenon is formalised. The relevant facts can be expressed as an obligatory inversion rule, if one posits an underlying SVO order for SS (Torrego 1984), or as a constraint on the preverbal position, assuming an underlying VOS order (Groos & Bok-Bennema 1986). The choice between these different representations does not affect the present analysis. The essential point is that the rule or constraint in question does not apply in CS.

As shown in Heap (1990), the most plausible explanation for this difference is that subject pronouns in Caribbean dialects are cliticised onto the verb, and are thus 'invisible' to the rule or constraint which applies in SS. Under this hypothesis, subject pronouns in this position are part of verbal morphology, and as such cease, in some sense, to be syntactic items. I am of course *not* claiming that subject pronouns *necessarily* undergo cliticisation: all the questions in (2) are also possible in CS. As Contreras (1989:175, note 12) says, "This does not mean that subject pronouns are always clitics in Caribbean Spanish. The claim is that the pan-Hispanic strong/weak alternation of non-subject pronouns has spread to subject pronouns in this dialect." I have nothing new to say here about the functional motivation of cliticisation of subject pronouns due to the aspiration and loss of the final *-s* of verbal morphology. Suffice it to note in passing that this functional pressure would only apply to the disambiguation of verbal forms in the second and third person singular, e.g. *(tú) tienes* in contradistinction to *(ella) tiene*.[2]

The cliticisation hypothesis therefore postulates a series of clitic subject pronouns which, although phonologically identical to the strong pronouns, can be distinguished from them syntactically by their appearance in preverbal position in *wh* questions. Although preverbal subject pronouns are presumably clitics throughout these dialects, I have chosen to examine this particular construction because it acts as a strong test: if subject pronouns can occur between a *wh* phrase and the verb in such questions, then they are to be considered clitics.

The published data leave a number of unresolved problems relating to the cliticisation hypothesis, two of which are discussed in the next section. New data from one Caribbean dialect are presented in section 2 and analyzed in section 3 in order to see how they shed light on these problems. I will conclude by discussing how the interplay between these phenomena illustrates the mutual dependence of morphological and syntactic change.

1.2. SOME PROBLEMS

Existing sources of documentation on this subject (reviewed in Heap 1990) do not agree on exactly which subject pronouns can appear in this construction. Some (Navarro Tomás 1948, Davis 1971, Quirk 1972) claim that only second person pronouns – *tú, usted, ustedes* – are possible, while others (cf. Heap 1990) attest to the occurrence of first and third person pronouns as well. While the cliticisation hypothesis does not predict whether all or only some subject pronouns can be cliticised,[3] what we may be dealing with here is a dynamic situation where cliticisation, once initiated by *tú*, is spreading to other subject pronouns.

Almost all authors seem to agree that lexical subjects cannot occur in the preverbal position characteristic of clitic pronouns. This is obviously consistent with the cliticisation

hypothesis, since the essential characteristic of clitic pronouns is that they "appear uniquely in positions that cannot be filled by an NP". (Kayne 1975:69) Thus if it is true, as some have suggested, that sentences like (3) can occur in CS dialects, the cliticisation hypothesis requires further elaboration, at the very least.

(3) ¿Qué Maria está haciendo? (Suñer, p.c.)
'what María is doing?'

Questions like (3) are of course ungrammatical in SS, and their grammaticality in CS is hereto undocumented. If they are indeed grammatical, their relationship (if any) to the occurrence of clitic pronouns in these dialects remains to be explained.

2. DATA FROM DOMINICAN SPANISH

In order to begin to answer these questions, I undertook a study of one CS dialect, that of the Dominican Republic (DR). Sentences elicited from four DR informants were compared to grammaticality judgements of a group of SS speakers. While judgements on both sides were at times far from unanimous, certain clear contrasts appeared.

2.1. SUBJECT CLITICS

The data in (4) clearly show that all subject pronouns – *nosotros, yo, tú, usted, ustedes, ella, él, ellas, ellos, uno* – can occur as clitics in DR.

(4) a. ¿Con quién nosotros nos encontramos ayer?
'with whom we (reflexive) met yesterday?'

b. ¿Cuántos innings yo piché?
'how-many innings I pitched?'

c. ¿De dónde usted viene?
'from where you (formal, singular) come?'

d. ¿Para qué ustedes vinieron?
'for what you (plural) came?'

e. ¿En qué tú trabajas?
'in what you (singular) work?'

f. ¿Con quiénes ellas hablan tanto?
'with whom (plural) they (feminine) speak so-much?'

g. ¿Qué ella le dijo cuándo él se enojó?
'what she to-him said when he (reflexive) got-angry?'

h. ¿Cómo uno puede llegar a la parada de guagua?

'how one can get to the bus stop?'

i. ¿Qué materia él enseña en la escuela?
'what subject he teaches in [the] school?'

j. ¿Qué ellos quieren comer?
'what they want to-eat?

Questions like these are uniformly judged ungrammatical by SS speakers, regardless of the pronoun involved. While DR speakers all accepted the data in (4) as grammatical, there was considerable variation as to the acceptability of the different subject pronouns. By far the most common subject clitic overall is *tú*, followed by *usted* and *ustedes*. Other pronouns, especially *nosotros*, were occasionally rejected in preverbal position by DR speakers, which suggests some basis for the belief that this construction only occurs with second person pronouns.

2.2 LEXICAL SUBJECTS

The data for non-pronominal subjects were less conclusive. On the whole, DR speakers tended to reject questions with preverbal lexical subjects, either changing the subjects into pronouns (as one might expect under the cliticisation hypothesis) or moving the subjects to postverbal position. Despite this tendency, the sentences in (5) were deemed grammatical in DR, and of course, completely ungrammatical in SS.

(5) a. ¿A quién el médico no visita nunca?
'whom the doctor *neg* visits never?'

b. ¿A quién Antonio no quiere visitar?
'whom Antonio *neg* wants to-visit?'

c. ¿Adónde esa gente va tan rápido?
'where those people go so fast?'

d. ¿Dónde tu mamá trabaja?
'where your mum works?'

The fact that far more lexical subjects in preverbal position were rejected than accepted is a strong indication that the cliticisation hypothesis is an essentially correct characterisation of the current state of DR. The acceptability of a limited number of lexical subjects in this construction suggests that cliticisation has triggered a syntactic reanalysis in this dialect. This development is explored in the next section.

3. ANALYSIS

The high frequency of second person pronouns as clitics can be attributed in part to pragmatic factors: questions with *tú* as a subject seem to be very common in general, so it is natural that they should appear frequently with clitic subjects in DR. It has been suggested that the monosyllabic nature of *tú* makes it a prime candidate for cliticisation, but this alone does not explain why it should be more common than *él* or *yo*.

I would not like, however, to suggest that the phonological length of the various pronouns is a crucial factor in cliticisation. It has been suggested that a three syllable pronoun such as *ustedes* or *nosotros* cannot be pronounced as a group with the verb, and therefore cannot be cliticised. Both the existing literature and data from this study indicate that this is not the case. What is more, sentences like (6) show that in French, where preverbal clitic pronouns are uncontroversially accepted, the cliticised group may have as many as four syllables.

(6) Pourquoi elle ne la leur a pas donnée?
'why she-*neg*-it-to-them has *neg* given?'

In short, phonological length does not seem to be the crucial criterion for what pronouns can become clitics in syntactic terms. In fact, data from this study show that in DR the preverbal clitic group may optionally include adverbs like the negative *no* and the aspectual *ya*, any subject pronoun, and object pronouns (direct, indirect or reflexive), as in (7).

(7) a. ¿Con quién ustedes ya no hablan más?
'with whom you (plural) no longer speak?'

b. ¿Qué tú no le dijiste?
'what you not to-him said?'

Let us stress here that, whatever the *phonological* status of this preverbal clitic group is, in *syntactic* terms the preverbal subject pronouns are exactly equivalent to the corresponding verbal morphology. When questioned as to whether there is any difference between forms like (8) a. and b., DR speakers consistently reply that they are 'the same thing' (i.e. neither is more 'basic'), in contradiction to questions like (8) c., which are considered an emphatic variant.

(8) a. ¿Qué quieres?
'what want (second singular)?'

b. ¿Qué tú quieres?
'what you (singular) want?'

c. ¿Qué quieres tú?
'what want YOU (singular)?'

It seems clear therefore that the two unmarked possibilities in DR correspond to syntactic structures like those in (9) a. and b., where the subject is recoverable without a strong pronoun.[4] (9)

(9) a. ¿Qué quieres?
$[_{CP}$ Qué$_i$ $[_{VP}$ $[_{V'}$ quieres $t_i]$ $e_j]]$

b. ¿Qué tú quieres?

$[_{CP}$ Qué$_i$ $[_{VP}$ [tú$_j$- $[_{V'}$ quieres $t_i]$ $e_j]$]]

It seems that the types of lexical subjects which have crept into preverbal position are those which, in some way, most resemble pronouns: simple NPs and proper names. Although my data on this point is still very preliminary, it is interesting to note that the NPs accepted in this position seem to be subject to some of the same conditions as clitics: they can be neither modified nor quantified, as shown by the contrasts in (10).

(10) a. ¿A quién el médico no visita nunca?
'whom the doctor *neg* visits never?'

b. * ¿A quién el nuevo médico no visita nunca?
'whom the new doctor *neg* visits never?'

c. * ¿A quién todos los médicos visitan?
'whom all the doctors visit?'

While it would be virtually meaningless to describe subjects like *el médico* in (10) a. as a 'cliticised NP', they are clearly 'closer' to pronouns in length or complexity than are modified or quantified NPs such as (10) b. and c.

4. CONCLUSIONS

It is very tempting to conclude that DR represents a transitional stage somewhat analogous to Middle French where, according to some authors,

> à mesure que les diverses personnes des verbes se confondaient davantage entre elles, les pronoms sujets se sont immobilisés devant le verbe en qualité d'affixes. (Bourciez 1956: 684)

The current facts, however, do not support such a teleological view. As we have seen, null subjects remain fully recoverable from postverbal morphology, and there is no reason to suppose that the introduction of clitic pronouns is changing this. Typologically, it can be argued that a grammar which uses both subject clitics and null subjects is highly unlikely (Roberge & Vinet 1989:58). If this were true, it might well be inherently unstable, and therefore evolve towards the elimination of one or the other (e.g. Modern French). Only continued observation of DR Spanish will tell us whether or not this is the case.

What we can learn from the current facts is how subject cliticisation may come about, and what other changes it may trigger. One important observation is that the class of subject pronouns does not homogeneously acquire the new morphosyntactic feature [+cliticisable] all at once. The differential frequency of cliticisation with the various subject pronouns suggests that this new feature has spread gradually from the pronoun *tú* where it is easiest to accommodate phonologically and pragmatically most likely, in a manner reminiscent of Chen and Wang's lexical diffusion model:

> '[it] seems apparent that this phenomenon [cliticisation], while evidently being rapidly generalized, spread from a nucleus of expressions involving *tú*, although by now the situation has become too confused for an accurate determination' (Lipski 1977:61).

While I do not have any direct evidence here to show where cliticisation began, it does appear to be most well-rooted with *tú*, and least well-established with *nosostros*. It

seems more than plausible that this synchronic variation reflects a diachronic change happening 'before our very eyes'.

Even before cliticisation spread throughout the whole subject pronoun paradigm, it in turn triggered a further change: the syntactic reanalysis which permits certain lexical NPs to appear as preverbal subjects. The motivation for this reanalysis may well be that, as the preverbal clitic group expands in the number and type of elements it can include, it superficially resembles a simple NP more and more closely.[5] This leads me to tentatively propose the following relative chronology as a likely sequence of changes leading up to the current state of DR Spanish:

(11) a. the strong/weak alternation spread from the object pronoun paradigm into the subject pronoun paradigm, thus allowing them to be cliticised, beginning with *tú*;

b. these subject pronouns form part of the preverbal clitic group, which may also include the negative *no* and adverbs such as *ya*, as well as object pronouns;

c. this clitic group is now sufficiently long in phonological terms that it is reinterpreted as equivalent to a simple NP subject;

d. this reinterpretation triggers a restructuring of question syntax in DR to allow for the order *wh* + SV with certain restricted types of lexical subjects.

Without further diachronic data, the sequence of events proposed in (11) obviously must remain very speculative. But if, as the synchronic facts suggest, it is essentially correct, then it shows how a morphological change like cliticisation can gradually spread through a particular category such as pronouns, and then lead to a syntactic change with possibly far-reaching consequences.

ACKNOWLEDGEMENTS

I would like to thank Yves Roberge, who directed the MA mémoire from which the data for this paper are drawn, and whose help has been invaluable. Terry Nadasdi, Parth Bhatt, Ed Burstynsky, and Jeff Tennant also provided many useful comments and suggestions. For both inspiration and encouragement, I cannot thank Claudia Vicencio enough. Naturally, only I am to blame for any remaining errors. Finally, I am grateful to the University of Toronto French Department for the financial support which allowed me to undertake and complete this research.

---

[1] This does not apply to yes/no questions, where the subject pronoun, if present, may appear before or after its verb.

[2] I have nothing to add regarding the proposal that cliticisation is due to contact of CS with either African substrata languages, or with adstrata languages such as English, an issue which is discussed fully in Heap (1990).

[3] Cross-linguistically, cliticisation can either affect the whole subject pronoun paradigm (as in French), or only part of it (as in Northern Italian dialects such as Trentino). See Roberge & Vinet (1989, chapter 2).

[4] Note that the representation given in (9) follows Contreras 1989 in assuming that the underlying order in Spanish is VOS. This choice is however in no way crucial to my analysis.

[5] Exactly *how* and *why* this reanalysis took place is a crucial question which I cannot answer as yet. While a number of deeper issues may be involved, it seems likely that the superficial identity of the strong and weak subject pronouns may have been a factor. A specific analysis of interrogative inversion in SS (a question which I have deliberately left open here) may also be necessary to account for this change.

REFERENCES

Bourciez, E. 1956. *Éléments de linguistique romane*, Paris: Klincksieck.

Burunat, S. et al. 1987. *El español y su sintaxis*. New York: Peter Lang Publishers.

Chen, M.Y. & S.-Y. Wang. 1975. 'Sound Change: Actuation and Implementation.' *Language* 51:2 255-281.

Contreras, H. 1989. 'Closed Domains.' *Probus* 1.2: 163-180.

Davis, J.C. 1971. 'Tú, ¿Qué tú tienes?' *Hispania* 54: 331-33

Groos, A. & Bok-Bennema. 1986. 'The Structure of the Sentence in Spanish.' in I. Bordelois, H. Contreras, K. Zagona (eds.) 1986, *Generative Studies in Spanish Syntax*. Dordrecht: Foris Publications, 67-80.

Harris, M. & Vincent, N. 1988. *The Romance Languages*, London: Croom Helm.

Heap, D. 1990a. 'Les questions à sujet pronominal antéposé dans les dialectes de l'espagnol des Caraïbes.' *Journal of the Atlantic Provinces Linguistic Association* 12: 13-38.

Heap, D. 1990b. *La Syntaxe des Questions à sujet antéposé en espagnol de la République Dominicaine: étude des cas en dialectologie contrastive* M.A. mémoire. University of Toronto..

Henriquez Ureña, P. 1975. *El español en Santo Domingo*. Santo Domingo: Editora Taller.

Kany, C.E. 1945. *American-Spanish Syntax*. Chicago: University of Chicago Press.(2nd edition: 1951.)

Kayne, R.S. 1975. *French Syntax: the Transformational Cycle*. Cambridge: MIT Press.

Lipski, J.M. 1977. 'Preposed Subject in Questions: Some Considerations.' *Hispania* 60: 61-67.

Navarro Tomas, T. 1948. *El español en Puerto Rico: contribución a la geografía lingüística hispanoaméricana.* Río Piedras: Editorial de la Universidad de Puerto Rico.

Quirk, R.J. 1972. 'On the Extent and Origin of Questions in the form ¿Qué tú tienes?' *Hispania*: 55: 303-304.

Roberge, Y. & M.T. Vinet. 1989. *La variation dialectale en grammaire universelle.* Montréal: Presses de l'Université de Montréal.

Torrego, E. 1984. 'On Inversion in Spanish and Some of its Effects.' *Linguistic Inquiry* 15.1: 103-129.

LEXICAL DIFFUSION AND WORD FREQUENCY IN PHONOLOGICAL BORROWING: *O/OU* AND *AIS/OIS* IN THE HISTORY OF FRENCH

Maurice Holder
St. Thomas University
Fredericton, New Brunswick

ABSTRACT

In the lexical diffusion of sound change, and of processes involving phoneme replacement such as analogy or borrowing, word frequency is often cited as a possible factor determining which lexical items are first affected. As regards dialect borrowing, it has been noted that the most frequent items are likely to be first affected. Two cases from the history of French are examined here: in one case, replacement of ou *by* o *borrowed from Latin, it is the least frequent words which are affected. This borrowing however falls into a different category of borrowing, namely* cultural borrowing *(Bloomfield, ch. 25), which by its very nature affects learnèd and literary (therefore infrequent) vocabulary first.*

The other case examined is a case of dialect mixture resulting in fluctuation between ais *and* ois *and eventual lexical redistribution of the two sounds. In one category, that of geographic adjectives, the two sounds are more or less evenly distributed among lexical items (e.g.* japonais/chinois*), and word frequency seems to be relevant only for a few items such as* anglais *and* français*. It is open to speculation what factors may have been responsible for the present-day lexical distribution of the two adjectival suffixes, and an ad hoc explanation will be attempted.*

1. INTRODUCTION

In recent years, considerable interest has been shown in the hypothesis of lexical diffusion of sound change, according to which some changes first affect only a few lexical items, then gradually spread to others – i.e. diffuse across the lexicon – until all relevant items are reached (Wang 1969, Chen and Wang 1975). Lexical diffusion is manifest in phonological borrowing, and the object of this paper is to reexamine two such cases in the history of French. To begin with, here is a brief review in four parts of some of the more important contributions in the literature on the subject: first, regular as opposed to lexically diffused sound change; second and third, other types of change involving sounds, namely analogy and borrowing; and fourth, the role of word frequency in these changes.

2. LEXICALLY REGULAR VS. LEXICALLY DIFFUSED SOUND CHANGE

The idea of lexical diffusion of sound change was first put forward by Schuchardt over a hundred years ago in opposition to the Neogrammarian regularity hypothesis of gradual sound change affecting all relevant vocabulary at once (note that Schuchardt did not actually use the term 'lexical diffusion'). The hundred year old controversy has been admirably summed up by Labov (1981). From his own surveys of sound change in progress, Labov presents evidence of different sorts supporting both sides of the controversy:

> We have located Neogrammarian regularity in low-level output rules and lexical diffusion in the redistribution of an abstract word class into other abstract classes. (304).

For example, low-level output rules producing vowel shifts *within* the subsystem of long vowels, were the raising of /ohr/ towards /uhr/ in New York (277), and the fronting of /ow/ and /uw/ in Philadelphia (283-4): *all* lexical items in these classes were affected. On the other hand, lexical diffusion seems to be operating in changes *across* subsystems. For instance, comparing the subsystems of lax and tense vowels in Philadelphia (286), Labov reports that words with short *a* such as *man, mad, had, glad, sad, dad, ran...*, have split into two classes, lax /æ/ with members *sad, dad, ran...*, and tense /æh/ realized [e:ə] to which *man, mad, had, glad...*, have been redistributed. The basis of these two polar types of sound change has been questioned by Betty Phillips (1984: 321-2) who gives examples in Old English and Middle English of lexical diffusion of changes *within* subsystems of short and long vowels. She states that such examples do suggest that Labov's two polar types may be oversimplified, though the existence of Neogrammarian sound change is not thereby disproved. Similarly, Wang (1979: 369) is of the opinion that "the truth of today will be the 'special case' of tomorrow ... The neogrammarian conception of language change will probably continue to be part of the truth." It seems that as more and more cases of lexical diffusion are uncovered, sound change in the Neogrammarian sense is being relegated to a narrower and narrower category of surface phonetic mechanics. As Schogt has observed (1962: 92): "Si la loi phonétique diachronique sans exception existe, son domaine est limité."

Consider now another kind of change operating alongside sound change.

3. SOUND CHANGE VS. PHONEME REPLACEMENT

3.1 ANALOGY

A clear line must be drawn between sound change and other mechanisms of change which might involve the replacement of one phoneme by another. The latter type of change is typically represented by phonological borrowing – see remark by Schogt in the next section. Also typical of this process is analogy, as when /ɛ/ in *crept* is replaced by /i:/ to form *creeped*, by analogy to the present tense form *creep* (Hooper 1976). This is not an example of sound change because there is no phonetic evidence elsewhere that /ɛ/ is changing in the direction of /i:/. The distinction seems to have been well appreciated by the Neogrammarians:

> it appears that what they were doing was motivating the choice between two competing regularities found in the course of comparative reconstruction. One was to be called sound change; the other analogy or dialect borrowing. (Hoenigswald 1978, quoted in Labov 1981: 272).

Hooper approaches the subject in much the same spirit:

> I will divide morpho-phonological change in the traditional way into sound change, which I view as largely, perhaps entirely phonetically motivated, and analogy, which I consider to be primarily conceptually motivated. (95).

She goes on to show the role of word frequency in lexical diffusion of both types of change in American English: this will be looked at in Section 4.

3.2 DIALECT BORROWING[1]

The other source of change often cited together with analogy is that of dialect borrowing. The latter subject is extensively covered in Bloomfield (1933), and a particularly interesting example from his book will be quoted below.

For examples from the history of French, one may turn to an article by Schogt (1962) where the author discusses under a number of headings the explanations usually given in historical grammars for exceptions to 'Neogrammarian' regular sound change. The headings are: purely phonetic factors, analogy, popular etymology, borrowing. Schogt illustrates the importance of dialect borrowing, **social** as well as **geographic**, in phonetic evolution. For instance, with regard to /ɛ ~ wɛ/ variation, which we will be looking at more closely in Sections 7 and 8 below, Schogt says (1962: 85, note 17): "Pour *e* les *phonèmes de remplacement* viennent de deux dialectes différents" (my emphasis). On the same subject, he points out the phonetically abrupt (*saut par emprunt*, 92), and lexically gradual character of borrowings (91): "cette substitution ne se fait pas simultanément dans tous les mots." In other words we have lexical diffusion, or, as summed up in the oft-quoted slogan, 'each word has its own history'. Finally, on innovation and conservation in different social classes, Schogt offers the following perspective:

> Passant maintenant à l'examen des couches sociales d'un seul dialecte géographique, nous constatons deux forces opposées: la langue populaire, riche en innovations, qui a pour elle le plus grand nombre, et la langue des classes aisées, qui est plus conservatrice et qui s'impose par son prestige. (91)

This view is supported by recent sociolinguistic research, according to Anthony Kroch (1978: 18, note 5). Its relevance will be seen below in the sections devoted to /ɛ~wɛ/ variation.

Turning to Bloomfield's treatment of dialect borrowing, among the many examples he gives, the following is probably as good a model as any (329-31): in Dutch dialect areas, one finds varying pronunciations of the words *mouse* and *house*; among these, figure the following in certain delimited areas: 1) [my:s, hy:s], 2) [mu:s, hy:s], 3) [mu:s, hu:s]. In Bloomfield's interpretation (361-2), one of the forms in 2) is due to borrowing. In the borrowing dialect, either the change to [y:] had not been made, and the form with [y:] was borrowed from a dialect where [y:] was the norm; or the change to [y:] had been made, and the form with [u:] was borrowed from a dialect that had not as yet developed [y:].

Borrowing as an explanation for exceptions to otherwise regular 'Neogrammarian' sound change, has been viewed with suspicion by lexical diffusionists, e.g. Chen 1977, Wang 1979. For instance in the case of tone changes in Chinese dialects, there is strong evidence pointing to system-internal developments rather than extralinguistic interference such as interdialectal borrowing (Chen 1977).

Chen however does not claim that dialect borrowing is simply ruled out in all cases. In his concluding remarks he states:

> The concept of time dimension is proposed as an explanation for exceptions to sound laws not to the exclusion of other theories, such as dialect mixture, but as an equally plausible hypothesis which deserves serious consideration. It goes without saying that whenever there is sufficient external evidence, e.g. historical account of considerable intermingling of speech communities or contemporary evidence of dialect geography, one may legitimately assume the presence of cross-

> linguistic or interdialectal interference. It is when such extralinguistic evidence is totally absent or dubious that the direct inference from irregular correspondences to dialect mixture or borrowing must be regarded as unwarranted and fortuitous. (245)

Bloomfield's example from Dutch posseses the ingredients listed by Chen in order to qualify as a genuine case of dialect mixture: it is drawn from a dialect geography study, and describes a situation resulting from the clash of two waves of cultural expansion, one from the [y:] pronouncing Frisian area, the other from the [u:] pronouncing North German Hanseatic cities; it is at the meeting place of these two cultural and dialectal currents in eastern Holland that there is dialect mixture, manifested in an etymological inconsistency, that of conservative [mu:s] existing side by side with advanced [hy:s] (see map in Bloomfield p. 328, from a dialect geography study by the Dutch dialectologist G.G. Kloeke). Bloomfield's rationale for borrowing is set out in the following passage:

> Our isoglosses of *mouse* and *house*, and doubtless many others, are results of the varying balance of these two cultural forces. Whoever was impressed by the Hollandish official or merchant, learned to speak [y:]; whoever saw his superiors in the Hanseatic upper class, retained the old [u:]. The part of the population which made no pretensions to elegance, must also have long retained the [u:], but in the course of time the [y:] filtered down even to this class. This process is still going on:...
>
> The word *house* will occur much oftener than the word *mouse* in official speech and in conversation with persons who represent the cultural center; *mouse* is more confined to homely and familiar situations. Accordingly, we find that the word *house* in the upper-class and central form with [y:] spread into districts where the word *mouse* has persisted in the old-fashioned form with [u:]. This shows us also that the Holland influence, and not the Hanseatic, was the innovator and agressor; if the reverse had been the case, we should find districts where *house* had [u:] and *mouse* had [y:]. (330-1)

In this scenario[2] the change takes place in the word *house* rather than *mouse*, because *house* is more frequently encountered in contact with the lending dialect. The role of word frequency is thus evoked, leading us into the next section.

4. THE ROLE OF WORD FREQUENCY

4.1 WORD FREQUENCY IN SOUND CHANGE AND IN ANALOGICAL CHANGE

Fidelholtz (1975) demonstrates that in English, vowels in frequently-used words reduce to schwa more often than in relatively rarely-used words. A related study on word frequency is that by Hooper 1976, quoted above. In this study Hooper confirms Schuchardt's observation in 1885 that sound changes affect the most frequent lexical items first, and that of Paul, around 1886, that analogical leveling tends to affect infrequent items first. She found that sound change, represented by schwa-deletion, affects the most frequent words first, and analogical leveling affects the least frequent words first (e.g. *celery, memory* with schwa-deletion, vs. *artillery, armory* without; *creeped, leaped, weeped*, with analogical leveling, vs. *kept, left, slept* without). Hooper reasons that since schwa-deletion usually occurs in casual and fast speech, and more frequent than infrequent words occur at this level of speech, it is not surprising that frequent words are first affected by this type of sound change. The reverse is the case with analogical change presumably because children will quickly and easily learn those suppletive forms which they often hear (e.g. *kept, left, slept*). However they will create analogical forms (e.g. *creeped, leaped, weeped*) for infrequent items whose suppletive forms have been less often heard and therefore less perfectly memorized (e.g. *crept, leapt, wept*): this is known as the 'imperfect learning hypothesis'.

In her conclusion, Hooper poses the question whether a similar relationship would hold between word frequency and sound changes other than non-reductive, e.g. vowel shifts. This question was taken up by Betty Phillips in a study alluded to above in connection with Labov's findings.

Phillips presents evidence for three sound changes in the history of English that have affected the least frequent words first. None of these was physiologically motivated, and a comparison with changes affecting the most frequent words first leads Phillips to make the following proposal which she calls the **Frequency Actuation Hypothesis** (336): "Physiologically motivated sound changes affect the most frequent words first; other sound changes affect the least frequent words first." Among the changes investigated by Phillips are two phonetically conditioned processes, one of which is physiologically motivated – raising of Old English /a/ followed by a nasal consonant; the other non-physiologically motivated – glide deletion in American English. Word frequencies correlate with these changes in the manner predicted by the hypothesis. Glide deletion, though phonetically conditioned in that it is subject to sequential constraints, is seen to have an analogic therefore non-physiologic source. Glide deletion in e.g. *tune, duke, news*, is seen as an analogical extension of the rule deleting the glide in e.g. *rude, lewd, sue, Zeus, chew, Jew.*

Another of the important studies which Phillips refers to in support of her hypothesis is Sherman's investigation (1975) of the lexical diffusion of noun-verb stress alternation in English from 1570 to 1800. The pattern in question is that which differentiates nouns from verbs by stress placement, nouns receiving initial stress, and verbs final stress, e.g. áffix/affíx. This pattern has been analogically extended to more and more noun-verb pairs over the centuries: from a mere three words attested in 1570 with this stress distinction, there has been a steady increase to well over 200 in present-day English, or approximately 15% of eligible noun-verb pairs. Phillips (p. 333) calculates that on average it is the less frequent pairs which have been affected.

4.2 WORD FREQUENCY IN PHONOLOGICAL BORROWING

The next question to be addressed is whether any relationship may exist between word frequency and phonological borrowing. As if in direct reply, an article by Marinel Gerritsen and Frank Jansen bears the title 'Word frequency and lexical diffusion in dialect borrowing and phonological change' (1980)[3]. The authors first of all carefully distinguish between sound change and phonological change brought about by dialect borrowing (32): "one has to assume in the latter case that the dialect variant is replaced by a variant from another dialect." (cf. section 2 above). They go on to comment on the relation between word frequency and attested dialect borrowing, in a passage which calls to mind the reasoning of Bloomfield (330-1 quoted above). Here are the comments in question:

> In the case of unnatural change, i.e. dialect borrowing, the following is to be expected. High frequency means that a word is often uttered, and consequently often heard. When people from Amsterdam consistently use the Amsterdam phonological variants in all words when speaking to dialect speakers, the latter are quite likely to hear the same frequent words with the Amsterdam variants, but far more rarely the less frequent words with the Amsterdam variants. In all probability, dialect speakers will hear and recognize the Amsterdam variants first in highly frequent words. It is acceptable, therefore, to assume that in the subsequent process of unconscious imitation of the new variant, the dialect speaker will use Amsterdam variants first and foremost in the words he frequently heard pronounced with the Amsterdam variants, thus in highly frequent words. This seems to be a sound reason for positing the following hypothesis:

> Hypothesis I: In dialects with an attested Amsterdam influence, i.e. an Amsterdam influence brought about by unnatural change, highly frequent words will develop towards the Amsterdam variant at a higher rate than less frequent words. (Gerritsen and Jansen 1980: 34).

Hypothesis I is confirmed in North Holland dialects neighbouring Amsterdam, where the authors found that the more frequent the lexical item, the higher the rate of change. In another study based on the same Dutch dialect material, Gerritsen and Jansen (1982) demonstrate, with the aid of a formula developed by Trudgill (1974), a procedure which helps to distinguish dialect borrowing from sound change.

Simply put, discreteness or phonetic abruptness has been proposed as the property which most typically distinguishes borrowed variants from sound change. For instance, in the conclusion to his article 'La notion de loi dans la phonétique historique' (1962), Schogt contrasts the phonetic gradualness of sound change with the abrupt leap taken in the process of borrowing:

> pour un petit noyau il s'agit d'une évolution graduelle de caractère phonétique aboutissant à un changement phonologique; le reste de la population connaît un saut par emprunt, sans être passé par les étapes phonétiques intermédiares. (92)

Similarly, John Harris in his recent work on phonological variation and change in Hiberno-English (1985), observes:

> One type of change which apparent-time evidence does show to involve discrete input and output categories is the redistribution of phonemes across the lexicon. Strictly speaking, this doesn't constitute phonological change proper (although it may eventually have phonological consequences). The phenomenon generally occurs in response to exonormative pressures and in traditional terms would be called borrowing. ... One of the things that quantitative analysis of phonological variation has allowed us to do is show clearly that the two types of change are qualitatively different and clearly distinguishable while they are in progress. Sound change proper is typically reflected in variation across phonetic continua; phonemic redistribution typically involves alternation between phonetically quite discrete variants. (2-3)

We are now ready to consider the two cases of phonological borrowing in the history of French announced in the title of this paper. In question are two pairs of formerly variant pronunciations where a great deal of fluctuation was noted before standardization was finally imposed. First, here is a word on written sources of information on these pronunciations.

5. DOCUMENTATION ON CHANGE IN PROGRESS IN THE HISTORY OF FRENCH

The first descriptions of the French language appear around 1520, as far as can be ascertained, and from that time on numerous treatises and compilations were published by grammarians. Fortunately, the material in a full 184 of these works, from the first quarter of the 16th to the middle of the 19th centuries, was gathered by Charles Thurot, and published in 1881-3 (vol. 1: vowels; vol. 2: consonants, nasal vowels, quantity, accent; vol. 3: index). Thurot sorted and classified the material into individual vowels and consonants according to phonetic environment, thus rendering an immense service to historical linguists. These descriptions afford us a much more direct insight into the state of the language than that which could be inferred in previous epochs from texts of literary or administrative character.

The cases of phonological borrowing about to be examined were already firmly entrenched in the 16th century in learnèd circles, but more or less resisted among the masses. The eventual result was lexical redistribution of the sounds within each vacillating pair: on the one hand [ɔ] and [u], (orthographically *o* and *ou*, e.g. *soleil/couleur*); on the other hand [ɛ] and [wɛ] (present-day [wa], orthographically *ais* and *ois*, e.g. *japonais/chinois*). I will be looking first at *o/ou*, which is the more straightforward of the two cases.

6. THE LEXICAL DISTRIBUTION OF *O/OU*

The normal outcome of Latin mid-high rounded vowels **ŏ, ō, ŭ** in pretonic position, was [u] (orthographically *ou*) in French . Thus we have *couronne* from **cŏrōna**, *couler* from **cōlare**, *tourment* from **tŏrmĕntu**, *souvent* from **sŭbĭnde**, and so on (Bourciez 1958). However, French was flooded with borrowings from Latin during the Renaissance period, and in such cases, pretonic [ɔ] remained unchanged, e.g. *collection, collège, colosse, copieux, objet, volume*, to cite only a few from many hundred, probably thousands. This was due initially to the restoration of Latin pronunciation, inaugurated by Erasmus (Fouché 1958: 427). By the 16th century, the newly introduced [ɔ] had gained much ground, and a host of words with *ou*, at all stylistic levels, were reported to have an alternative pronunciation with *o* (Thurot vol 1: 252-66). After a period of fluctuation, one of the forms of each word was standardized, and the other fell into disuse. For example, doublets such as *soleil/souleil*, *brossailles/broussailles*, which abounded in the 16th century and after, did not survive in double form, and *soleil* and *broussailles* were standardized at the expense of *souleil* and *brossailles*.

Among the many words where fluctuation between [ɔ] and [u] was actually noted and commented on by the grammarians, the pattern of standardardization which eventually emerged was as follows: forms with [ɔ] usually ousted those with [u] in words of learnèd or literary character, whereas the reverse was true for more familiar, everyday words. As for direct borrowings from Latin (i.e. *latinisms*) these had [ɔ] only, and no fluctuation with [u] was ever attested (this is the case with words such as *collection, collège* etc. noted above). Once the sound [ɔ] had been identified with *latinisms*, the pattern was set for *relatinisation* of words from the native stock by **replacement** of [u] with [ɔ]. The main candidates for *relatinisation* of this sort were words which fell within the learnèd, literary, and administrative sphere, but the innovation spread to familiar vocabulary as well. It was eventually rejected at this level no doubt because of its association with the upper layer of vocabulary (a few discrepancies remain, see below).

The distribution of the two sounds correlates rather well with word frequency, since the learnèd and literary terms which accepted the change to [ɔ] are by and large less frequent than the everyday terms which did not. This proportion is the reverse of that found for dialect borrowing by Gerritsen and Jansen. The reason is evident: the borrowing of Latin [ɔ] falls into a different category of borrowing, namely *cultural borrowing* (Bloomfield, ch. 25), which by its very nature would affect infrequent words first.

In addition to borrowing, there are other factors, phonetic and analogical, which are considered by historians of the French language to have played a role in the assignment of these vowels to different words. These categories will now be considered with the help of examples in Table 1. This table gives a sample of words whose pretonic vowel fluctuated between [ɔ] and [u] in the 16th, 17th, and 18th centuries, as attested in the writings of the grammarians of that time (Thurot: ibid.). The words are given in present-day form, and

divided into five categories as follows: *phonetically conditioned o* (+ nasal consonant), *analogical o*, *borrowed o* (learnèd words), *borrowed o* (familiar words), *phonetically regular ou* (familiar words).

PHONETICALLY CONDITIONED *O*	ANALOGICAL *O*	BORROWED *O* (LEARNÈD)	BORROWED *O* (FAMILIAR)	PHONETICALLY REGULAR *OU* (FAMILIAR)
fromage	côté	forêt	orgueil	mourir
froment	fossé	forfait	ortie	nourrir
homard	ôter	projet	corvée	pouvoir
romarin	poser	portrait	rosée	vouloir
commander	cochon	colonne	rossignol	tourner
commencer	poteau	volonté	colombe	douleur
concombre	corbeau	volume	soleil	couleur
connaître	corbeille			courage
honneur	corneille			couronne
bonnet	ormeau			courroie
tonner	broder			moulin
donner	folie			mouton
sonner	mollesse			fourmi
raisonner	mortel			pourceau
étonner	porter			tourment
trognon	prochain			broussailles
				souvent
				souvenir

Table 1.
Sample of words in present-day French with historically attested fluctuation between *o* and *ou* forms.

In the category of *phonetically conditioned* change, nasalization and subsequent lowering are said to be responsible for the vowel *o* in such words as *fromage, froment, commencer* etc. There are in fact very few words with *ou* in this category (*poumon* seems to have undergone the influence of the Latin spelling). Abundant testimony can be found in Thurot (vol. 2: 511-25) of an earlier pronunciation of *o* + nasal as *ou*.

The words in the second column presumably underwent the *analogical* influence of forms with [ɔ], the regular outcome for Latin **ŏ** in a **stressed** closed syllable. Thus, analogical *porter* and *prochain* fluctuated with phonetically regular *pourter* and *prouchain* in Middle French; but the forms with *ou* were ousted by the forms with *o*; the latter, *porter* and *prochain*, were created by analogy with the phonetically regular *porte* and *proche*. Old French *coste* (Mod. *côte*), Old & Middle French *coche, post, corb*, etc. are said to have furnished the model for *côté, cocher, poteau, corbeau* etc. The very common verbs *pouvoir* and *vouloir* are not however influenced by *peux* and *veux*, nor is *mourir* influenced by *meurs* or *mort*: this lends support to the hypothesis that the most frequent words would not be affected by analogical change (see Section 4).

In the *learnèd* category, only forms with *o* borrowed from Latin subsist. The inclusion of a word such as *forêt* in this category may seem surprising; Bourciez (101) states that the form with *o* was adopted contrary to the rule, but etymological research in fact shows juridical influence here (Bloch and von Wartburg 1964).

Familiar words with borrowed *o* form a minor anomalous group of words which, it is said, were exposed to literary influences (even so, forms such as *souleil* and *rousée* were

the norm in works by the poets of the Pléiade). In the category of *familiar* words, it is evidently the phonetically regular *ou* which predominates.

There was also fluctuation between *o* and *ou* in the South-western dialects like Poitevin and Saintongeais, a state of affairs continued in Franco-Acadian with forms such as *oblier, jornée, pomon, moton, houmard, coument, coumencer, estoumac, counaître, douner.*

The next change to be looked at is somewhat more complex, and does not follow the criterion of word frequency in such a clear-cut manner. It has been scrutinized in all the historical grammars, presents many a thorny issue, and has also been the object of a doctoral thesis by Schogt (1960), on which work I rely for most of the background given in Section 7. I then focus on word frequency in Section 8.

7. THE LEXICAL DISTRIBUTION OF *-AIS/-OIS*

The historical facts regarding this split in French are briefly as follows: Latin tonic free ɛ̄ and ɪ̄, i.e. in a stressed open syllable, diphthongized to *ei* and then followed two divergent paths: on the one hand simplification to *è* (normal spelling *ai*); on the other, differentiation of the first element of the diphthong resulting in *oi*, opening of the second element giving *oè* and *oa*, finally stress shift to the second element yielding *wè* and *wa* (normal spelling *oi*) (according to some historians *ai* is a simplification of *wè* rather than *ei*, but this has little bearing on the question under consideration).

The monophthong *ai* was a characteristic feature of the western dialects, Norman and Breton in particular; the evolution to *oi* characterized the eastern dialects. Paris was the meeting place of these two dialectal currents coming from either direction, east and west. In terms of general phonetic evolution, Parisian speech was western in character (Schogt 1960: 78). The forms *dret, fret, crère* in vernacular French Canadian (= *droit, froid, croire*), go back no doubt to French western dialect forms.

By the 16th century when the first descriptions of the French language appeared, *oi* (= [wɛ] with popular variant [wa]) had already spread to most relevant lexical items in the region of Paris. There were however two large categories where *-ais* (= [ɛ]) subsisted: the imperfect/conditional ending, and geographic adjectives. We know from the remarks of the grammarians from the 16th century onwards that in these categories *-ois* was favoured by the educated classes, whereas *-ais* was distinctly working class Parisian (Thurot vol 1: 374-414, Beaulieux 1967: 170-4, Schogt 1960).

Two factors help to explain why *-ais* was accepted: first, an important verbal category does not react in the same way as isolated substantives (Schogt 1960: 114). In other words, *-ais* was more stable and resistant to variation because of its status as a morphological marker.

Secondly, and paradoxically, *-ais* came into style in courtly circles, supposedly because many of the courtiers from the provinces wished to appear Parisian by adopting this typically Parisian pronounciation (Schogt 1960: 101). There can hardly be a doubt that the prestige of the court was a positive influence leading to the eventual adoption of *-ais* as the norm for the imperfect ending and for many geographic adjectives. We therefore have two sub-types of dialect borrowing, each occurring in a different social class: in courtly circles, **social** dialect borrowing of Parisian vernacular *-ais*; in educated circles, **geographical** dialect borrowing of eastern *-ois*. In a curious reversal of roles, the

originally innovative *-ois* had become a mark of conservative educated speech, whereas the originally conservative *-ais* was now making the rounds as an aristocratic novelty. There was subsequent lexical redistribution of the two sounds, resulting in the present-day lexical distribution as follows (examples in Table 2):

1) *ai* only, for the imperfect and conditional: three paradigmatic endings *-ais, -ait, -aient.*

2) both *ai* and *oi* for geographic adjectives: masculine endings *-ais, -ois*; feminine: *-aise, -oise.*

3) *oi* in most other cases: most nouns, pronouns, non-geographic adjectives, and verbs: infinitive endings *-oir, -oire*, and conjugated forms *sois, doit, crois, vois,* etc.

The trajectory we can trace, as far as the historical records permit, for diffusion across the French lexicon of *oi* as a geographic dialect borrowing, is as follows: in educated speech, the innovative *oi* spread to all relevant lexical items, then after a period of fluctuation with *-ais*, the ending *-ois* was completely ousted as marker of the imperfect tense (category 1), but survived as a geographic adjective-forming suffix sharing the field with *-ais* (category 2). The fact that geographic adjectives were formed with the aid of a variety of suffixes (*-an, -ain, -ien, -éen, -ais*) meant that the system could easily accommodate an extra variant in the form of *-ois*. The opposite tendency towards leveling of variants was manifest in the case of the imperfect tense.

In category 3) *oi* was dominant, but there are a certain number of exceptions where *ai* was standardized after a period of fluctuation with *oi* (Table 2): in nouns with the ending *-aie* from Latin **-ēca**, **-ēta**, a few non-geographic adjectives, infinitives *connaître* and *paraître*, and before a nasal consonant (*moins, foin*, and *avoine* are exceptions to the exception! – see Bourciez 1958).

Among the 'petit peuple', the situation was similar to that just outlined, except that the imperfect seems never to have been pronounced *oi* at all. This is so according to one of the most respected historians of the French language, Charles Beaulieu (1967):

> N'y a-t-il pas lieu de penser que la diphtongue *ei*, dans un certain nombre de substantifs et les formes verbales de l'imparfait et du conditionnel, avait subsisté dans le langage populaire, aussi bien à Paris qu'en province, tandis qu'au Palais, et partout où on parlait un langage soutenu, *oi* s'était toujours substitué à *ei*? ... mais ... dans le XIV^{e} siècle, *oi* a gagné tous les mots; ce qui ne veut pas dire que le peuple ait cessé de prononcer *èi, è*. Il nous paraît au contraire infiniment vraisemblable que, dans les cas cités plus haut, jamais le peuple ne prononça *wè* et que c'est au contraire la bonne société qui a fini par prononcer comme le peuple. (vol. 1: 171)

The patterning of *ai* and *oi* according to grammatical category is clear, but it is mainly in category 2), geographical adjectives, that a possible connection with word frequency may be sought. This is dealt with in the next section.

IMPERFECT ENDING	NOUNS	GEOGRAPHICAL ADJECTIVES	GEOGRAPHICAL ADJECTIVES	INFINITIVES	NOUNS
-ais	monnaie	anglais	chinois	voir	toile
-ait	craie	français	danois	avoir	soir
-aient	taie	écossais	finnois	savoir	soie
	claie	irlandais	gallois	devoir	mois
CONDITIONAL	raie	hollandais	gaulois	pleuvoir	voie
	aunaie	finlandais	hongrois	recevoir	poil
-rais	roseraie	groenlandais	suédois	pouvoir	poire
-rait	verre	islandais	québécois	vouloir	poivre
-raient	tonnerre	zélandais	niçois	boire	loi
		portugais	viennois	croire	foi
		polonais	berlinois	croître	etc.
		japonais	bruxellois		
		montréalais	hambourgeois		
		new-yorkais	strasbourgeois		
	INFINITIVES	OTHER ADJECTIVES	OTHER ADJECTIVES	PRONOUNS	
	connaître	faible	bourgeois	moi, toi	
	paraître	raide	courtois	soi, quoi	
		frais	étroit		
		épais	froid	NUMERAL	
			droit	trois	
		NASAL	NASAL		
		frein	foin		
		plein	moins		
		veine	avoine		

Table 2.
Sample of the lexical distribution in present-day French of [ɛ] and [wa] from Latin tonic free ĕ, ĭ.

8. GEOGRAPHIC ADJECTIVES

It has been seen that the pronunciation *-ais*, popular in courtly circles, was borrowed from the Parisian vernacular, whence it found its way indirectly into standard usage by way of the Palace. It has also been seen (Sections 3 and 4) that in dialect borrowing, one may expect the most frequent words to be affected first. In accordance with this prediction, one should therefore expect the borrowed suffix *-ais* to be used for the most frequently mentioned nations, whereas the less frequent would more likely receive the suffix *-ois* associated with educated usage.

Frequency of usage is in fact specifically linked by 18th century grammarians with the selection of the *-ais* suffix over *-ois* (Thurot vol. 1: 402). Buffier for instance, in the year 1709, declares: "*François, Anglois, Ecossois, Irlandois, Hollandois, Polonois*, se prononcent communément en *è*, *Francès, Anglès, Ecossès*; mais les noms des nations avec qui nous avons moins de raport *[sic]* se prononcent en *oè*: *Suedoès, Danoès, Siamoès, Hongroès*." A similar remark is made by Duclos in 1754: "Si nous étions dans une relation aussi habituêle d'afaires, de guère, et de comerce avec les Suedois et les Danois qu'avec les Anglois, nous prononcerions bientôt *Danès* et *Suedès*, comme nous disons *Anglès*." Frequency is considered a possible factor by well-known historians of the French language such as Ferdinand Brunot (1966: 541), Albert Dauzat (1967: 115, note 1) and Pierre Fouché (1958). Fouché states (277) that the ending *-ais* was employed for the most

frequently used geographical adjectives, but subsequently he does qualify this considerably.

Anglais and *français* may indeed be more frequent than *suédois* and *danois*, but once we get beyond these, things are not so clear-cut, and purely phonetic factors may be at work: notice that countries ending in *-lande* all take the ending *-ais*: now it is conceivable that countries like Holland, and perhaps Ireland were mentioned with some degree of frequency, but Finland, Greenland, Iceland and New Zealand hardly so. It may be then, that once the suffix *-ais* was used for countries ending in *-lande* such as *Irlande* and *Hollande*, giving the adjectives *irlandais, hollandais*, then this served as a rhyming model for all other countries ending in *-lande*, **regardless** of frequency: *finlandais, groenlandais, néerlandais, islandais, zélandais*. This interpretation is I feel reinforced by the fact that the ending *-andois* seems rare in French. It is not listed in Alphonse Juilland's *Dictionnaire inverse de la langue française* (1965), and the only example I have come across (purely by chance I may add) is *Dinandois* for inhabitants of the town *Dinan* in France (note the triplet *Dinannais/Dinandois/Dinandiens*, and doublets *français/François, l'anglais/Langlois, Monréal/Mont Royal*, etc. – a resource well exploited by the French language). As regards rhyming patterns, other possibilities which might be considered are: *polonais - japonais, chinois - finnois - berlinois*.

A subjective factor of taste and euphony may also be involved, leading people to prefer the *-ois* ending for *québécois*, but *-ais* for *montréalais*: this does seem to be a matter for speculation in our present state of knowledge. Another example where we may speculate that the choice of suffix was perhaps guided by euphony is the case of the adjectives *fransasquois* and *saskatchewanais*. In the latter, a subjectively comical repetition of [wa] is avoided by the choice of *-ais*, whereas in the former, the preceding [k] seems (subjectively again) to make *-ois* the natural choice.

9. CONCLUSION

Two cases of phonological borrowing in the history of French have been presented. In one case, that of the cultural borrowing of pretonic [ɔ] from Latin, there can be no doubt as to the direction and limits of lexical diffusion, and the manner in which it can be correlated with word frequency. The borrowing above all affected learnèd, and therefore less frequent vocabulary, then filtered down to more frequent everyday items. A major portion of relevant vocabulary was reached, since fluctuation between *o* and *ou* was abundantly attested in the 16th century. Eventually however *o* was rejected in favour of *ou* for the majority of familiar, everyday items.

The other case presented, that of dialect mixture resulting in assignment of *-ais* to some lexical items, and *-ois* to others, remains problematic. Assignment by grammatical category provides a partial explanation of a general kind, namely that different grammatical categories react differently to a phonological innovation. It is mainly in one category, that of geographic adjectives, that both forms are found. Assignment of one form or the other to a given adjective seems arbitrary for the most part. Though frequency was actually mentioned by 18thcentury grammarians, on close examination this factor only seems relevant for a few of the commonest adjectives such as *anglais* and *français*. Purely phonetic factors may also be at work, but here the speculative element has the upper hand. We are not much further ahead on this matter and there things must rest for the time being. It remains to be seen whether new approaches may help to explain some of the old problems. Modern research on dialects in contact (Trudgill 1986) provides some interesting

new viewpoints, such as accommodation theory, which would certainly be worth following up.

---

1 I wish to thank Sarah G. Thomason, keynote speaker at this conference, for suggesting that I go further into this question.

2 Since Bloomfield's main source of information here was Kloeke's study, I am not certain to what extent Bloomfield's ideas echo those of Kloeke – see Gerritsen and Jansen 1982: 12.

3 With respect to titles, it was difficult for me to avoid 'borrowing' from the title of Gerritsen and Jansen's article, since the theoretical subject matter they treat is identical to that which is presented here.

REFERENCES

BEAULIEUX, CHARLES. 1967. *Histoire de l'orthographe française. Tome Premier: Formation de l'orthographe des origines au milieu du XVI^e siècle*. Nouveau Tirage. Paris: Champion.

BLOCH, OSCAR and WALTHER von WARTBURG. 1964. *Dictionnaire étymologique de la langue française*. Quatrième édition. Paris: Presses Universitaires de France.

BLOOMFIELD, LEONARD. 1933. *Language*. New York: Holt.

BOURCIEZ, ÉDOUARD. 1958. *Précis historique de phonétique française*. Neuvième Édition. Paris: Klincksieck.

BRUNOT, FERDINAND. 1966. *Histoire de la langue française des origines à nos jours. Tome I: De l'époque latine à la Renaissance*. Nouvelle édition. Paris: Colin.

CHEN, MATTHEW. 1977. 'The time dimension: Contribution toward a theory of sound change.' In WANG (ed.) 1977: 197-251. (Reprinted from *Foundations of Language* **8** (1972), 457-498).

CHEN, MATTHEW and WILLIAM S-Y. WANG. 1975. 'Sound change: Actuation and implementation.' *Language* **51**. 255-81.

DAUZAT, ALBERT. 1967. *Tableau de la langue française*. Paris: Payot. (Première édition 1939).

FIDELHOLTZ, JAMES L. 1975. 'Word frequency and vowel reduction in English.' *Papers from the Eleventh Regional Meeting*, Chicago Linguistic Society. 200-13.

FOUCHÉ, PIERRE. 1958. *Phonétique historique du français. Volume II. Les Voyelles*. Paris: Klincksieck.

GERRITSEN, MARINEL and FRANK JANSEN. 1980. 'Word frequency and lexical diffusion in dialect borrowing and phonological change.' *Dutch Studies* **4**. 31- 54. The Hague: Martinus Nijhoff.

_____. 1982. 'The interplay between diachronic linguistics and dialectology: Some refinements of Trudgill's formula.' *Papers from the 3rd International Conference on Historical Linguistics*, ed. by J. Peter Maher, Allan R. Bomhard and E.F. Konrad Koerner. 11-37. Amsterdam: John Benjamins.

HARRIS, JOHN. 1985. *Phonological variation and change: Studies in Hiberno-English.* Cambridge: Cambridge University Press.

HOENIGSWALD, HENRY M. 1978. 'The Annus Mirabilis 1876 and posterity.' *Transactions of the Philological Society.* 17-35.

HOOPER, JOAN B. 1976. 'Word frequency in lexical diffusion and the source of morphophonological change.' *Current progress in historical linguistics. Proceedings of the Second International Conference on Historical linguistics*, Tucson, Arizona. 95-105. Amsterdam: North-Holland.

JUILLAND, ALPHONSE. 1965. *Dictionnaire inverse de la langue française.* The Hague: Mouton.

KROCH, ANTHONY. S. 1978. 'Toward a theory of social dialect variation.' *Language in Society* **7**. 17-36.

LABOV, WILLIAM. 1981. 'Resolving the neogrammarian controversy.' *Language* **57**. 267-308.

PHILLIPS, BETTY S. 1984. 'Word frequency and the actuation of sound change.' *Language* **60**. 320-42.

SCHOGT, HENRY G. 1960. *Les causes de la double issue de E fermé tonique libre en français.* Amsterdam: van Oorschot.

_____. 1962. 'La notion de loi dans la phonétique historique.' *Lingua* **10**. 79-92.

SHERMAN, DONALD. 1975.' Noun-verb stress alternation: An example of the lexical diffusion of sound change in English.' *Linguistics* **159**. 43-71.

THUROT, CHARLES. 1881-3. *De la prononciation française depuis le commencement du XVI^e siècle, d'après les témoignages des grammairiens.* Three volumes. Paris: Imprimerie nationale. (Slatkine Reprints, Geneva, 1967).

TRUDGILL, PETER. 1974. 'Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography.' *Language in Society* **2**. 215-46.

_____. 1986. *Dialects in contact.* Oxford: Blackwell.

WANG, WILLIAM S-Y. 1969. 'Competing sound changes as a cause of residue.' *Language* **45**.9-25.

_____. 1977. (ed.) *The lexicon in phonological change.* The Hague: Mouton.

_____. 1979. 'Language change – A lexical perspective.' *Annual Review of Anthropology* **8**. 353-71.

LA GRAMMAIRE TEXTUELLE DE WEINRICH

Jean-Marc Lemelin
Université Memorial

RÉSUMÉ

Harald Weinrich inscrit sa grammaire dans la linguistique textuelle et dans une anthropologie de la communication et il la définit comme étant dialogique, interactionnelle et instructionnelle; le fondement de cette grammaire est sémantique plutôt que syntaxique. C'est ainsi qu'une trentaine de traits pertinents sémantiques permettent de structurer la morpho-syntaxe des catégories grammaticales, de manière à faire apparaître le primat de la jonction textuelle sur la fonction phrastique. En outre, le caractère fondamentalement proprioceptif plutôt que cognitif du sémantique distingue cette grammaire textuelle des autres grammaires textuelles et de la grammaire cognitive. Enfin, elle se démarque autant de la grammaire générative que de la grammaire générale.

Le développement de la linguistique est évidemment inséparable de celui de la grammaire: même si la linguistique du début du XX^{e} siècele s'est démarquée, d'une part de la grammaire philosophique, qu'elle soit «raisonnée» ou autre, et de la grammaire comparée, d'autre part de la grammaire normative ou corrective et de la philologie, il demeure que la grammaire descriptive et explicative est le tronc, sinon la racine, de la linguistique. On pourrait même aller jusqu'à dire – et de plus en plus – que la linguistique n'est rien d'autre que la grammaire. Mais tout dépend alors de ce que l'on entend par *grammaire*. Ce qui est ici présupposé, c'est que la grammaire est un ensemble de règles, de lois ou de principes, un ensemble de contraintes par lesquelles est possible la créativité; en ses sous-composantes phonologique et morphologique, syntaxique et sémantique, la grammaire est la principale composante de la *langue*, l'autre composante étant le *vocabulaire*, sans toutefois aborder ici le problème du rapport entre la grammaire et le *lexique* et sans traiter du statut de la lexicologie et de la lexicographie par rapport à la linguistique, le dictionnaire – le répertoire ou la mémoire linguistique – étant déjà une grammaire mais la grammaire étant irréductible au dictionnaire... Objet de la linguistique ou linguistique elle-même, la grammaire peut avoir une *dominante* morphologique, syntaxique ou sémantique; ce peut être une grammaire du mot, de la phrase ou du texte. La grammaire de Harald Weinrich, qui est ici présentée, est une grammaire textuelle[1].

1. LINGUISTIQUE TEXTUELLE ET ANTHROPOLOGIE DE LA COMMUNICATION

Weinrich situe sa grammaire dans une linguistique textuelle et dans une anthropologie de la communication [p. 19]; c'est-à-dire que le corpus qui est son objet n'est ni le mot ni même la phrase, mais le *texte* défini comme: «l'énoncé linéaire qui est compris entre deux interruptions remarquables de la communication et qui va des organes de la parole ou de l'écriture de l'émetteur aux organes de l'audition ou de la vue du récepteur [p. 24]». Les textes les plus authentiques sont les textes dialogués, qu'ils soient parlés ou écrits. Dans le *dialogue*, il y a priorité de la perspective de l'auditeur ou du récepteur, «car seuls les messages qui atteignent leur destinataire ont leur pleine signification

communicative [p. 20]». Caractéristique de la situation de communication, référence et forme fondamentale de la communication, le dialogue oral en face-à-face, le jeu dialogal, met en action le *corps*, qui est le système premier du langage [p. 20]; c'est-à-dire que les données corporelles de la communication sont immanentes à la *situation*: cette *immanence*, c'est en quelque sorte *l'énonciation* présupposée par *l'énoncé* qu'est le texte. De la «situation environnante» ou de l'«environnement situationnel», de ce que l'on pourrait appeler le *site*, fait partie le *contexte* (ou l'entourage verbal), qui ne doit pas être confondu avec le *référent*, l'autre aspect du site.

La linguistique textuelle de Weinrich n'est pas seulement dialogique ou dialogale et interactionelle, elle est aussi, sinon surtout, instructionnelle. La grammaire étant «ce système de règles méta-communicatives permettant un usage efficace des textes [p. 26]», sont nécessaires une série *d'instructions* ou d'impératifs, c'est-à-dire de «consignes qu'un émetteur adresse à un récepteur pour lui faire savoir comment on attend de lui qu'il se comporte dans la situation donnée [p. 20]». Ces instructions sont fournies par une trentaine de «traits pertinents sémantiques élémentaires qui se présentent par paire et en opposition binaire dans chaque paire [p. 21]»; chaque instruction implique une contre-instruction et cette opposition est neutralisable dans un trait sémantique neutre. Ces traits constituent «le fondement universel ou quasi-universel que le français a en commun avec d'autres langues [p. 21]» et ils permettent de construire tous les concepts syntaxiques, un *concept grammatical* se définissant «comme la combinaison particulière d'un nombre limité de traits pertinents sémantiques [p. 21]». C'est donc dire que la grammaire de Weinrich est essentiellement sémantique: la sémantique y fonde la syntaxe et la syntaxe, la morphologie; et, dans cette grammaire, «les traits sémantiques grammaticaux jouent le rôle de fondement des traits lexicaux [p. 636]», c'est-à-dire des traits sémantiques dans les *lexèmes*; en outre, entre les consignes que sont les *morphèmes* et les lexèmes, il y a parfois une certaine influence réciproque. En fin de compte, les *significations* sont des instructions réussies, un texte ayant du *sens* «quand le récepteur peut régler son comportement là-dessus [p. 25]»; c'est pourquoi le texte est irréductible aux signes, qui sont les «plus petits segments significatifs d'un texte [p. 25]», mais qui n'ont pas nécessairement de sens au niveau global d'un texte ou indépendamment de celui-ci. En d'autres mots, la *surface* du signe, donc le phonème (sans signification, selon Weinrich), est débordée par la *face* ou *l'interface* de la signification, sans laquelle il n'y a pas de sens, c'est-à-dire de *volume* et donc de *profondeur*...

2. LA MORPHO-SYNTAXE DES CATÉGORIES GRAMMATICALES

À partir de ces «fondements méthodologiques» et de ces «concepts linguistiques fondamentaux», Weinrich procède à l'élaboration de sa morpho-syntaxe des diverses catégories grammaticales: l'accord (impliquant entre autres le nom), les rôles textuels, le verbe, l'article, l'adjectif, l'adverbe, la jonction et le dialogue. Plutôt que d'insister, comme le font la plupart des grammaires phrastiques, sur le nom et le verbe ou sur la proposition, Weinrich est amené à privilégier les *particules*, soit la classe fermée des morphèmes plutôt que la classe ouverte des lexèmes dans leur comportement grammatical, la syntaxe y étant ainsi une affaire de *jonctions* encore plus que de *fonctions*. Au niveau des *rôles textuels*, sont distingués les «communicants» ou les «rôles communicatifs» et les «actants» ou les «rôles actanciels». Les *communicants* correspondent aux personnes grammaticales de la situation de communication; ce sont les actants de la communication ou de l'énonciation:

- première personne: *locuteur* (ou scripteur),

- deuxième personne: *auditeur* (ou lecteur),
- troisième personne: *référent.*

Les *actants* proprement dits sont les actants de la narration ou de l'énoncé:

- origine de l'action: *sujet*,
- co-actant: *partenaire*,
- *objet* de l'actance [p. 58].

Mais «dans la langue française les communicants et les actants sont toujours amalgamés [p. 59]», la *conjugaison* étant l'indication des rôles textuels dans le verbe et d'autres indications se retrouvant dans le nom et le pronom ou dans un morphème présentatif ou un «morphème d'horizon» comme le 'il' dit impersonnel [p. 80]. Les rôles textuels permettent, entre autres choses, de diviser la situation de communiction en l'*arrière-plan* et l'*avant-plan*; c'est par les rôles textuels qu'il y a, avant toute *référence*, amorce du *repérage* contextuel (le débrayage anaphorique) et du repérage situationnel (l'embrayage déictique), pourrait-on ajouter, comme principale «instruction de recherche»...

Fidèle à un de ses ouvrages antérieurs[2], Weinrich assimile la syntaxe du verbe à la syntaxe des temps grammaticaux, le *temps* se caractérisant par sa perspective (rétrospective/prospective), son registre (commentaire/récit ou documentaire) et son relief (premier plan ou avant-plan/arrière-plan) [p. 117]; ce sont ces critères qui commandent la conjugaison de tous les verbes jusque dans la composition. La *perspective* temporelle concerne la distinction entre le temps du texte et le temps de l'action [p. 121]; le *registre* conditionne l'attitude de réception, le commentaire conseillant ou commandant l'attente, l'attention, voire la tension, alors que le récit permet la détente ou la distance et laisse donc plus de place à l'imagination [p. 124]; le *relief* est la focalisation du thème de l'arrière-plan, où il n'y a que topicalisation, par le rhème de l'avant-plan. La mise en relief est ainsi substituée à *l'aspect* et au «mode d'action»[3]. Quant aux *modes* comme l'impératif et le subjonctif, ils sont d'abord et avant tout analysés comme instructions données à l'auditeur, instructions lui permettant de s'orienter dans le texte ou dans la situaiton [p. 169], ou comme engagement de la part du locuteur, engagement signalé par divers *déclencheurs* [p. 176-177].

Les modalités sont rapidement abordées, un peu de la même manière, et reléguées à l'arrière-plan dans les récits [p. 191]. Les morphèmes nominaux ou les prédéterminants que sont les articles comprennent les articles simples et les articles spécifiques. L'*article simple* peut être anaphorique (défini) et ainsi renvoyer à une pré-information (connue): c'est un «signal de routine ('thématique') [p. 206]»; l'article simple peut aussi être cataphorique (indéfini) et alors renvoyer à une post-information (inconnue): c'est «l'annonce d'un élément nouveau ('rhématique') et un signal de vigilance [p. 206-207]». L'article partitif est rendu possible par la neutralisation de l'opposition de nombre que seul l'article cataphorique permet [p. 202]; quant à l'article-zéro (ou neutre), il se retrouve fréquemment avec les noms propres [p. 201]. L'*article spécifique* peut être démonstratif, possessif, numéral, sommaire (indéfini) et interrogatif ou exclamatif. Beaucoup de *pronoms* ne sont finalement que les formes libres de l'article [p. 230] ou ils peuvent être considérés comme étant des morphèmes de focalisation ou des morphèmes présentatifs [p. 237].

La troisième classe de lexèmes, l'*adjectif*, a comme trait sémantique la détermination de la base (déterminable) par un *qualificatif* (déterminant) [p. 271]; l'attribut est considéré comme prédication adjectivale et l'épithète comme qualification adjectivale [p. 272]. Ce qui préoccupe surtout Weinrich à propos de l'adjectif, c'est sa position: par la place postposée ou post-déterminante, il y a une plus grande force de détermination que par la place antéposée ou pré-déterminante, où l'adjectif n'apporte pas alors tous ses traits lexicaux au nom et il a donc le statut de (quasi-)morphème [p. 273-274]. C'est ainsi que dans l'antéposition de l'adjectif, par exemple, «par sa position dans le texte, un signe lexical (lexème) est transformé en signe grammatical (quasi-morphème) [p. 279]». Ces adjectifs antéposés peuvent être des morphèmes d'assertion, de diminution et d'augmentation, de rôles, d'articles sommaires ou des morphèmes temporels [p. 280 et ss.]. A l'inverse, l'article numéral en post-position se trouve de cette manière lexématisé en quasi-adjectif [p. 287]. Avec les adjectifs, les *adverbes* sont des qualificatifs, mais ils déterminent au contraire une base non-nominale [p. 316]. La classe des adverbes est composée de mophèmes adverbiaux, de lexèmes adverbiaux et de locutions adverbiales mixtes [p. 317-319]. S'il est topicalisé, l'adverbe a une détermination faible; s'il est focalisé, il a une détermination plus forte; la focalisation vient généralement de la dernière position «qui est la plus expressive du point de vue sémantique: on y loge en particulier l'information nouvelle ('rhématique') [p. 320]». La postposition de l'adverbe est donc commandée par sa longueur et sa signification forte [p. 320]. Weinrich fait aussi remarquer que les adverbes de position «ont en principe à la fois une composante spatiale et une composante temporelle [p. 330]». Enfin, les adverbes temporels sont, comme les verbes, définis en termes de perspective, de registre et de relief [p. 337 et ss.]; les adverbes sommaires sont les formes libres de l'article sommaire [p. 339] et il n'y a pas de morphèmes adverbiaux dans la classe des adverbes de qualité [p. 396], mais surtout des lexèmes adverbiaux formés à partir d'adjectifs à signification qualitative et non classifiante [p. 397].

3. VERS UNE GRAMMAIRE PROPRIOCEPTIVE?

C'est dans la syntaxe de la jonction et dans la syntaxe du dialogue que la grammaire de Weinrich s'avère être une grammaire vraiment textuelle. La *jonction* est une «liaison textuelle», un «réseau de détermination qui est régi sémantiquement par un JONCTEUR» qui «signifie à l'auditeur dans quel sens il doit envisager la détermination» entre la base (déterminable) et le *complément* (déterminant) [p. 359]. Le *joncteur* est un morphème individuel qui oriente la détermination mais qui est inséparable de l'ensemble de la jonction et de la position du complément. Weinrich distingue ici les *joncteurs simples* comme 'et' et 'ou', les *prépositions* (qui lient la base et un complément nominal ou pronominal ou qui ont un infinitif pour complément), les *conjonctions* (qui lient une base verbale et un complément verbal) et les *relatifs* (qui lient une base nominale ou pronominale et un complément verbal et qui se rangent entre les prépositions et les conjonctions): ce sont des morphèmes ou des lexèmes morphématisés [p. 360].

Ce qu'il s'agit de retenir, surtout au sujet des prépositions – et en cela, la grammaire textuelle de Weinrich se rapproche, mais se distingue, de la grammaire cognitive[4] – c'est que «les prépositions de l'orientation tirent leurs significations d'un modèle anthropologique reproduisant les conditions corporelles d'une communication en face-à-face [p. 375]», mais non de la géométrie, d'une géométrie cognitive de la raison par exemple. «C'est pourquoi les significations de ces prépositions peuvent relever de certains organes de communication et de leur fonction caractéristique [p. 375]»; le champ de vision (qui inclut la bouche et les oreilles), l'orientation verticale ou latérale par «l'asymétrie communicative des mains [p. 382]» et la portée permettent de diviser ces prépositions. Il

n'y a aucun doute ici que Weinrich touche à une grammaire proprioceptive du corps, grammaire tributaire de l'imagination, qui schématise, et non de la raison, qui formalise. Weinrich ajoute:

> On peut mesurer les distances: c'est l'affaire de la géométrie. Les procédés de mesure propres à la géométrie ne sont pas pertinents pour les prépositions. La mesure des prépositions est fournie par les organes de communication et leur fonctionnalité. Certaines prépositions définissent ainsi leurs significations en fonction de la portée des organes de communication, laquelle est en principe limitée. Dans la langue parlée naturelle ces limites sont fixées, en particulier, par la portée de la voix et celle de l'oreille. Mais elles peuvent toutefois être élargies par la communication écrite ou amplifiée par des moyens électroniques [p. 383-384].

Les prépositions de portée ont donc, elles aussi, «une signification corporelle communicative et non une signification spatiale et géométrique [p. 386]».

La grammaire textuelle de Weinrich ne saurait pour autant être asssimilée à une simple pragmatique, ni même à l'anthropologie fonctionnelle dont il se réclame pourtant. La situation de communication dont il parle si souvent est d'abord et avant tout une situation d'énonciation, c'est-à-dire de signification, où le corps des communicants est l'*ancrage* du langage des actants: la *proprioceptivité* (thymique et tensive), qui est à la base de l'extéroceptivité (pragmatique) et de l'intéroceptivité (cognitive), n'est pas d'essence transcendante mais immanente; l'immanence de la signification est l'essence de la transcendance de la communication – et non l'inverse. L'objectif ou la destination de la signification, qui est la communication ou l'action, ne doit pas être confondu avec son trajet ou sa trajectoire, qui est l'énonciation. Mais de l'énonciation présupposée par l'énoncé, l'énonciation qui est elle-même énoncée, part de l'énoncé, n'est jamais que l'*étayage*... Plutôt donc qu'une *pragmatique*, au delà de la grammaire, la linguistique textuelle de Weinrich est davantage une *grammatique*, en deçà de la grammaire.

Cette grammaire textuelle est certes une grammaire *énonciative*, comme celle de Culioli ou de Joly et d'autres[5], mais ce n'est pas simplement une grammaire *discursive*, plus proche de la rhétorique et de la poétique, comme la narratologie de Genette[6], qui est issue d'une autre linguistique de l'énonciation[7]. Le texte n'est pas seulement le discours ou la performance, car il ne serait alors que le contexte (verbal); or, ce qui ressort même de la syntaxe du dialogue, c'est que la situation d'énonciation n'est pas ce qui s'ajoute au contexte de l'énoncé mais bien ce à quoi s'ajoute l'énoncé: bien avant le message, il y a le *contact* phatique et pathique, emphatique et empathique; contact des corps et jonction par des formules de salutation ou d'apostrophe, par des morphèmes phatiques, par des interjections [p. 490]; avant toute actance, il y a la prégnance de la *chair* des interlocuteurs. S'il y a une *génétique* du langage, voire de la langue, ce ne peut être qu'à ce niveau proprioceptif et non cognitif. Si «le dialogue est la structure-canon de la communication [p. 489]», c'est que le contact (thymique, proprioceptif) est la structure-canon du dialogue, celui-ci serait-il encore un monologue. Le contact est en quelque sorte l'*in*compétence sans laquelle il n'y a aucune compétence, l'incompétence intransitive du langage rendant possible la compétence transitive de la langue, dont le discours est la saillance (ou la discontinuité de la performance). La grammaire énonciative de Weinrich pourrait donc être située entre la grammaire cognitive, avec laquelle elle partage une certaine *schématique* de l'imagination[8], et la grammaire *sémio-narrative* qu'est la sémiotique de Greimas[9], sur le point de soumettre la logique de la raison et l'ontologique de l'action à une *agonistique* de la passion. En résumé, la grammaire de Weinrich n'est pas textuelle faute d'être phrastique: elle ne serait alors que *péri* ou *para*phrastique et donc *interprétative*[10]; elle est textuelle parce qu'elle est davantage phrastique, la phrase n'étant pas que cette relation sujet/prédicat mais étant déjà texte, textualité narratique et rythmique du récit[11] et donc de la continuité: *ad*verbalisation par la jonction plus que verbalisation par la fonction. Autrement que

comme *phase* de l'énonciation, que comme l'énoncé qu'est le texte, la phrase n'existe pas – ou n'existe que comme exemple écrit!

4. GRAMMAIRE UNIVERSELLE OU GRAMMAIRE GÉNÉRALE?

Tout au long de son ouvrage, Weinrich ne manque pas d'illustrer son propos par toutes sortes d'analyses de textes parlés ou écrits, courts ou plus longs, afin de faire saisir la *textualité* d'un texte, c'est-à-dire «la cohérence particulière qui fait qu'un texte est un texte»; c'est selon lui un objectif prioritaire de la description linguistique [p. 25]. Pour la linguistique textuelle, le texte-en-situation est la donnée première, dont il faut d'abord décrire le contexte linguistique et puis expliquer la situation, qui n'est pas que linguistique et qui est plus complexe. Pour y parvenir, Weinrich s'en remet toujours, en dernière instance, aux trente traits pertinents sémantiques ayant un caractère quasi-universel; caractère qu'il précise ainsi:

> Nous employons ici le terme de 'quasi-universel', parce que ces traits sémantiques peuvent servir à la construction des concepts grammaticaux d'autres langues, mais ne fournissent pas nécessairement la base grammaticale de toute langue. On doit cependant admettre que les concepts fondamentaux de toutes les langues peuvent s'élaborer à partir d'un nombre relativement restreint de traits sémantiques élémentaires, ce qui nous fonde à considérer ceux-ci comme le matériau de construction premier de cette grammaire [p. 635].

Nul doute alors que la grammaire textuelle de Weinrich est une grammaire profonde, mais où la profondeur est sémantique et non syntaxique, comme chez Greimas et contrairement à Chomsky; mais ce n'est cependant pas une grammaire universelle comme la grammaire générative. Est-ce pour autant une grammaire générale, comme celle de Pottier[12]? Si oui, ce n'est certes pas une grammaire *comparée* en ce qu'elle se veut synchronique, même si elle ne méprise pas «la dimension historique de la language [p. 22]». D'une certaine manière, il est possible de dire que c'est une grammaire *intégrée*, dans cette tentative d'en arriver à une théorie de *la* langue à partir d'*une* langue, le français, et dans cette entreprise qui a pour cadre une véritable grammaire *fondamentale*, fondamentalement sémantique.

Il importe de souligner ou de rappeler une dernière fois que, dans le sémantique, les catégories proprioceptives prévalent sur les catégories cognitives; les traits sémantiques concernent le rapport entre la personne, l'espace et le temps; rapport qui ne peut être synthétisé ou ponctué – c'est la *deixis* – que par l'imagination. Ainsi le sémantique, au fondement non sémantique du sens ou à son non-fondement, a-t-il un statut non pas pragmatique ou cognitif mais proprioceptif, c'est-à-dire thymique ou contactuel: déictique. Ce qui veut dire que le sens est autrement sémantique que la sémantique ne l'a jusqu'ici proposé, ou qu'il ne l'est pas dans les termes de la sémantique. Pour la grammaire fondamentale ou proprioceptive qu'est en fin de compte la grammaire textuelle de Weinrich, le sens ne saurait être *génératif*: c'est la signification qui l'est; le sens, lui, est *génitif* et *gérondif*: à l'origine ou à la racine, il est toujours en train de se faire. Une grammaire *radicale* serait une telle grammaire génitive et gérondive et elle pourrait se réclamer de cette science proprioceptive qu'est la science ordinaire ou générale de l'homme. Cette *pragammatique* du sens comme monde et langage est à l'écoute de la langue comme théorie intrinsèque ou immanente de l'univers ou comme esthétique transcendantale[13].

* * * * * * * * * *

Pour terminer sur quelque chose d'un peu moins abstrait, il n'est pas inutile de se demander en quoi la grammaire du français de Weinrich peut contribuer à l'enseignement,

si on considère qu'un *programme* de recherche n'est rien sans un *diagramme* ou un dispositif de réflexion par lequel il y a projet et trajet de *découverte*, c'est-à-dire d'*ouverture* à/de la pensée. D'une part, la grammaire textuelle s'inscrit dans une *théorie* du texte qui a déjà fait ses preuves en études littéraires, dans ce qu'il faut considérer comme étant une approche radicalement immanente des textes; approche pour laquelle l'objet n'est pas imposé par le corpus mais (pro)posé par la méthode, celle-ci jouant le rôle d'un 'syntagme' permettant de clore le 'paradigme'. D'autre part, dès le début de son ouvrage, Weinrich se montre soucieux de la *pratique*, de la didactique des langues et plus particulièrement de l'enseignement du français langue étrangère. Il ne manque pas, à chaque fois que l'occasion se présente, de donner quelques trucs pédagogiques et d'examiner, par exemple, le rôle que la *mémoire* peut jouer dans l'apprentissage. Cependant, sa grammaire intervient, dans l'*acquisition*, davantage au niveau de la spécialisation qu'au niveau de l'apprentissage proprement dit. Traduite en français, elle convient déjà mieux à des étudiants francophones d'origine et les autres étudiants, à moins d'avoir une connaissance fort avancée du français, ont très certainement besoin de l'intermédiaire d'un enseignant enthousiaste et compétent en linguistique. Cela n'empêche pas que cette grammaire puisse être un instrument pédagogique important pour ceux qui croient que la linguistique doit être un guide pour la didactique des langues et qu'une meilleure connaissance de la langue et des mécanismes d'acquisition de la langue maternelle est indispensable à l'enseignement des opérations d'apprentissage d'une langue seconde.

On ne sait ce que l'on enseigne que si l'on enseigne ce que l'on sait; en d'autres mots, la *manière* d'enseigner (ou l'expression) importe peu si la *matière* (ou le contenu) que l'on enseigne est encore beaucoup trop informe: telle pourrait être la premiée ou la derniée leçon à tirer de cette lecture (dia)grammatique de *Grammaire textuelle du français.*

* * * * * * * * * *

P.S. Lors de la discussion, il a surtout été question de la **valence** du verbe, c'est-à-dire de sa puissance d'attraction ou de répulsion de l'objet selon le nombre d'actants qu'il peut admettre, la valence étant caractéristique de la situation actantielle. La monovalence est le fondement de la plurivalence: tous les verbes peuvent en principe être monovalents; le modèle en est le verbe 'être', qui n'admet qu'un actant, le sujet, comme tous les verbes intransitifs. Il y a aussi les verbes qui admettent l'actant qu'est le partenaire: la bivalence sujet/partenaire caractérise d'autres verbes intransitifs (ou transitifs indirects). La bivalence sujet/objet est évidemment ce qui caractérise les verbes transitifs (directs), dont le verbe 'avoir' est le modèle. La trivalence sujet/partenaire/objet réunit les deux bivalences et est la situation de la co-actance. En outre, même si la valence est une donnée lexicale (et qu'il peut y avoir un dictionnaire de valences), elle peut avoir une réalisation grammaticale ou textuelle différente de ce code: en *survalence*, un actant objet peut être ajouté à un verbe monovalent ('Il vit sa vie', 'Il pleut des cordes', 'Il pleure des larmes'); en *sous-valence*, un actant objet ou partenaire peut être supprimé. Le passif est une sorte de sous-valence; la reflexivité est aussi tributaire des rôles textuels que sont les actants et les communicants [p. 58-116].

NOTES

1. Harald Weinrich. *Grammaire textuelle du français.* Traduction de l'allemand par Gilbert Dalgalian et Daniel Malbert. Didier/Hatier (Alliance française). Paris; 1989 [1982] (672 p.). Toutes les références entre crochets renvoient à ce seul ouvrage.

2. Harald Weinrich. *Le temps; le récit et le commentaire.* Traduction de l'allemand par Michèle Lacoste. Seuil (Poétique). Paris; 1973 [1964] (336 p.)

3. Voir *Le temps*, p. 107-130, ainsi que David Cohen. *L'aspect verbal.* Presses Universitaires de France (Linguistique nouvelle). Paris; 1989 (272 p.)

4. Ronald W. Langacker. *Foundations of Cognitive Grammar; Volume 1: Theoretical Prerequisites.* Stanford University Press. Stanford: 1987 (XII + 516 p.) et George Lakoff. *Women, Fire and Dangerous Things. What Categories Reveal about the Mind.* The University of Chicago Press. Chicago-London; 1987 (XVIII + 624 p.)

5. Antoine Culioli. *Quelques articles sur la théorie des opérations énonciatives.* Paris; 1965-1983 (160 pages photocopiées). André Joly. *Essais de systématique énonciative.* Presses Universitaires de Lille (Psychomécanique du langage). Lille; 1987 (336 p.). Jean Cervoni. *L'énonciation.* Presses Universitaires de France (Linguistique nouvelle). Paris; 1987 (128 p.). Jacqueline Guillemin-Flescher. *Syntaxe comparé du français et de l'anglais; problèmes de traduction.* Ophrys. Paris; 1981 (552 p.). A. Grésillon, J.L. Lebrave et al. *La langue au ras du texte.* Presses Universitaires de Lille (Linguistique). Lille; 1984 (208 p.)

6. Gérard Genette. *Figures III.* Seuil (Poétique). Paris; 1972 (288 p.) et *Nouveau discours du récit.* Seuil (Poétique). Paris; 1983 (128 p.)

7. Roman Jakobson. *Essais de linguistique générale.* Minuit (Points #17). Paris; 1963 (256 p.). Émile Benveniste. *Problèmes de linguistique générale 1 et 2.* Gallimard nrf (Tel #7 et 47). Paris; 1966 et 1974 (VIII + 360 p. et 294 p.). Catherine Kerbrat-Orecchioni. *L'énonciation; de la subjectivité dans le langage.* Armand Colin (Linguistique). Paris; 1980 (288 p.)

8. Mark Johnson. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason.* The University of Chicago Press. Chicago-London; 1987 (XXXVIII + 238 p.)

9. A.J. Greimas et J. Courtés. *Sémiotique; dictionnaire raisonné de la théorie du langage.* Hachette Université (Langue, linguistique, communication). Paris; 1979 et 1986 (2 tomes de 426 p. et 272 p.)

10. Pour une sémantique interprétative susceptible d'échapper à ce reproche, voir. François Rastier. *Sémantique interprétative.* Presses Universitaires de France (Formes sémiotiques). Paris; 1987 (II + 286 p.) et *Sens et textualité.* Hachette Supérieur (Langue, linguistique, communication). Paris; 1989 (288 p.)

11. Pour une analyse du récit plus propre à intéresser les linguistes, à part Weinrich lui-même et Greimas, voir: Jean-Michel Adam. *Le texte narratif; traité d'analyse textuelle des récits.* Nathan (Linguistique française). Paris; 1985 (240 p.). Jean-Michel Adam et André Petitjean. *Le texte descriptif; poétique historique et linguistique textuelle.* Nathan (Linguistique française). Paris; 1989 (240 p.). Laurent Danon-Boileau. *Produire le fictif; linguistique et écriture romanesque.* Klincksieck. Paris; 1982 (184 p.). Marcel Vuillaume. *Grammaire temporelle des récits.* Minuit (Propositions). Paris; 1990 (128 p.)

12. En partie à partir de la psychomécanique de Gustave Guillaume, Bernard Pottier. *Linguistique générale; théorie et description.* Klincksieck (Initiation à la linguistique, série B; Problèmes et Méthodes, 3). Paris; 1985 [1974] (340 p.)

13. Qu'il nous soit permis ici, en guise de justification, de renvoyer à notre dernier ouvrage: *Oeuvre de chair; de l'âme et du corps.* Ponctuation/Triptyque. Montréal; 1990 (132 p.)

CROSS-CATEGORIAL FORMAL IDENTITY: A FUNCTIONAL ACCOUNT

Cheng Luo
University of Manitoba

ABSTRACT

Linguistic change in morphology and syntax can involve inter-categorial transfer of the same form, often reflected synchronically through formal identity among coexisting categories. Such transfer can be motivated and conditioned by functional / semantic similarity between the categories involved. This study shows formal identity between the contrastive focus marker, the demonstrative and the copula in five languages: Mandarin, Hebrew, Margi, Mokilese and Kusaiean, where the three categories are shown to be by and large diachronically related. This relationship appears to be motivated by their sharing of the same functional domain of deixis.

1. INTRODUCTION

In many languages, the cleft construction resembles that of an equational sentence, with the configuration

$$[_S \text{ <Cpl> } X''_i \text{ <Cpl> } [_{S'} \text{ (Comp) } [_S \ldots [_{X''_i}\ e] \ldots]]]$$

where the contrastive focus occurs initially[1] and adjacent to a form that is identical to a copula:

ENGLISH

(1) It is *the cat* that killed the bird.

IRISH (McCloskey 1979)

(2) Is *é* *Seán* aL thigeann 'na bhaile
cpl him John that come home

'It's John that comes home.'

BRETON (Timm 1987)

(3) *un* *hunvre* *eo* *ho* peus bet.
a dream cpl 2pl-have had

'It's a dream that you had.'

While it is true that contrastive focus in these languages is marked mainly by the structural configuration, the adjacent copula also helps mark the focus. Without the copula, a cleft sentence without a Comp heading the presupposed clause, as would be in (3), would be indistinguishable from other constructions such as one with a topicalized NP. The copula thus used functions as a contrastive focus marker(CFM).

The focus-marking function of the copula-like form is more clearly seen in languages where the contrastive focus occurs in the same position as in a corresponding non-cleft sentence:

MANDARIN

(4) wo *shi* *zuotian* kanjian ta de.
I cpl/CFM yesterday see he MM
'It was yesterday that I saw him.'

MOPUN (Frajzyngier 1987)

(5) n-kwat *à* *sii* *siwol.*
1sg-pay cpl/CFM with money
'It is with money that I paid.'

The copula-like particle (*shi* in (4) and *à* in (5)) thus used is different from those in (1-3) in that the focus remains an argument within a one-clause structure. Consequently, the CFM is much less verb-like than in languages such as English and French. Crucially, the cleft reading would be lost without the presence of the copula-like CFM.

There are still other languages where the form of the CFM appears to be associated more with that of a demonstrative pronoun than of a copula, as in (6-7).

HEBREW

(6) *ze* *ani* she-pihakti.
dem/CFM I who yawned
'It is I who yawned.'

MARGI (Hoffmann 1963)

(7) *ni* *-n* atsianyi.
I this kill-him
'It is I who killed him.'

As will be shown shortly, some of these languages also exhibit formal identity between the CFM, the copula and the demonstrative pronoun in one way or another. Given this formal identity, is there any functional motivation that underlies these three grammatical categories?

This paper will address the issue by examining the presumably close formal relationship between the CFM, the copula and the demonstrative, in an attempt to explore a unitary functional account of such relationship. The theoretical framework within which the study has been done is basically functional, with special reference to some iconicity principles such as the isomorphism principle proposed in Haiman (1985), and the principle of semantic similarity which motivates cross-categorial formal identity as discussed in Stein (1988). It will be shown that linguistic change in the morpho-syntax of a language can involve transfer of the same form from one category to another, transfer which is often motivated and conditioned by semantico-functional similarity between the linguistic categories involved. Such formal identity is often reflected synchronically through formal

identity among coexisting categories. Since the claim that cross-categorial formal identity can be motivated by semantico-functional similarity will be supported and gain potential universality if the same phenomenon is found crosslinguistically, especially in unrelated languages, this study brings under examination data mainly from five languages: Mandarin Chinese, Hebrew, Margi, Mokilese and Kusaiean, each, as will be shown, in some way exhibiting identity in form between the three grammatical categories aforementioned.

2. THE MANDARIN CASE

As shown in (4), the CFM in modern Mandarin is *shi*, which is identical in form with the copula used in equational sentences with nominal predicates:

(8) Zhang San shi ta-de didi.
Zhang San cpl he-Gen younger-brother
'Zhang San is his younger brother.'

This same form was used as a demonstrative in Archaic Chinese(11th c. B.C. - 3rd c. B.C.). Given that the CFM, the copula verb and the demonstrative all use the same form, there are at least three possible explanations for this homogeneity:

(9) a. it is just a coincidence;

b. they may be diachronically related to each other;

c. they may have a common pragmatic function.

(9a) is untenable before we have had to reject (9b) and (9c). In what follows I will discuss (9b) and (9c) in turn. I will begin with the following hypotheses:

A. The diachronical change for *shi* is: demonstrative > copula > CFM.

B. The demonstrative, the copula and the CFM all share a common pragmatic function.

2.1 SYNCHRONIC EVIDENCE

2.1.1 Hypothesis A presupposes some diachronical relationship between the demonstrative, the copula and the CFM, and suggests that there is a closer diachronical relationship between the demonstrative *shi* and the copula *shi*. One piece of synchronic evidence in support of this claim is that in Mandarin, the copula *shi* is in general subject to the same structural constraint as the demonstrative pronoun *shi* in its determiner use, i.e. normally it can only occur before an NP (cf. Hashimoto 1969), whereas the CFM *shi* can occur before non-NPs as well as NPs. Compare the following examples:

(10) a. ta shi xuesheng (cpl NP)
he cpl student
'He is a student.'

b. *ta shi hao / chi fan / cong Zhongguo. (*cpl AP/VP/PP)
he cpl kind eat meal from China
'He is kind/eating/from China.'

(11) a. *shi* *wo* zuotian zai jie-shang kanjian ta de. (CFM NP)
CFM I yesterday Loc street-on see he MM
'It is I who saw him on the street yesterday.'

b. wo zuotian *shi* *zai jie-shang* kanjian ta de. (CFM PP)
'It is on the street that I saw him yesterday.'

In (10), *shi* is used as a copula, which is normally not stressed and should occur before nominal predicates. (10b) is problematic due to the presence of the copula before a non-NP[2]. In (11), however, *shi* is used as a CFM, which can appear before non-NPs as well as NPs, usually with some additional stress on the focused NP or PP or on itself if the focus is an AP or VP.

2.1.2 There are some apparent counterexamples to the constraint that the copula *shi* appears only before NPs. First, consider (12), where *shi* appears before an AP/VP and is stressed:

(12) a. ta **shi** (hen) hao.
he cpl very kind
'He is indeed (very) kind.'

b. ta **shi** lai-guo Jianada.
he cpl come-Perf Canada
'It is true that he's been to Canada.'

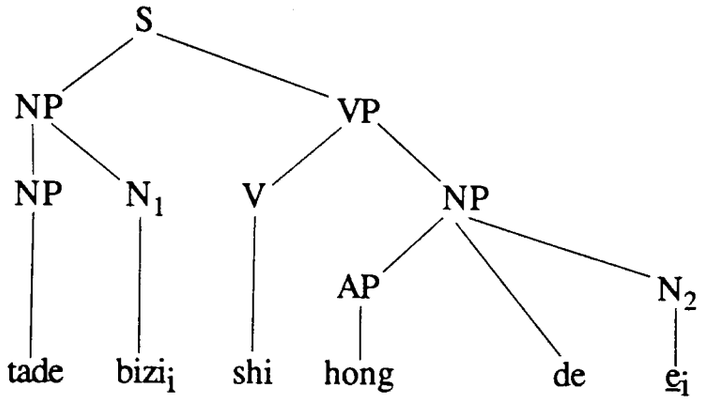
If *shi* were the copula, the fact that it appears before an AP and a VP respectively in the grammatical sentences (12a) and (12b), would be a violation of the constraint. However, I contend that *shi* should be analyzed not as copula but as CFM (cf. EMPH marker in Chomsky 1957) for the following reasons. First, (12a-b) are good only if *shi* is stressed (cf. (10b)), in which case the positive polarity of the predicate is contrasted with its negative polarity. For example, (12b) would be an appropriate negative response to the assertion 'he never came to Canada', with the implication 'it is not that he's never been to Canada'. This contrastiveness is clearly a feature associated with the CFM rather than with the copula. Secondly, the copula *shi* in Mandarin functions as the main verb in a sentence by virtue of being grammatically obligatory normally (eg. **ta xuesheng* 'he (is) a student'); while the CFM *shi* functions as a quasi-verb (Luo, ms.) or an adverb (Huang 1982), and is grammatically optional, as would be the case with (11), where the absence of *shi...(de)* would not affect grammaticality. Since (12a-b) are still grammatical without *shi*, it must be a CFM rather than a copula.

Next, consider (13), where *shi* is not stressed:

(13) ta-de bizi shi hong de.
he-Gen nose cpl red MM
'His nose is red.'

Structurally, *shi* in (13) occurs before an AP, a seeming violation of the pre-NP constraint. However, I will argue that (13) actually has an underlying structure like (14):

(14)



where the unrealized N_2 is coreferential with N_1. In support of this analysis is the fact that the AP cannot occur independently after the copula without the following *de* (cf. (10b)):

(15) *tade bizi shi hong.

If the copula could indeed occur before an AP, we should expect (15) to be grammatical, which is in fact not the case. On the other hand, this unstressed *shi* cannot be a CFM, which when occurring before an AP/VP should be stressed (cf. (12)). As for *de*, it is recognized as a general-purpose modifier marker in Mandarin (Ross, 1983), which may occur with or without a following head noun, especially when the latter is a noun with a general meaning such as 'person' or 'thing', or when its meaning is recoverable from the context, as is the case here. The obligatory cooccurrence of *de* with *shi* in (13), then, suggests that what underlyingly follows *de* is an unrealized head noun (='bizi'). Thus what looks like a *copula AP* structure has in fact an underlying *copula NP* structure. The constraint that the copula, like the demonstrative, occurs only before an NP is respected.

2.1.3 Synchronic evidence presented above shows that in modern Mandarin, the copula *shi* is in general subject to the same constraint as the demonstrative *shi* (in Archaic Chinese), whereas the CFM *shi* is free of this constraint. Considering this in historical context, if we tentatively assume that the demonstrative is older, there will be three logically possible routes of derivation:

(16) a. demonstrative > CFM > copula;

b. demonstrative > copula > CFM;

c. demonstrative > { copula, CFM }

The argument presented so far favours (16b), which is theoretically supported by Lightfoot's (1979) Transparency Principle, which 'requires derivations to be minimally complex and initial, underlying structures to be 'close' to their respective surface structures' in order to preserve communicability between generations. New categories are created through reanalysis of grammatical constructions whereby exceptionality is grammaticalized and derivational complexity eliminated. In other words, changes in grammar may make the existing initial structure analyses perceptually opaque to the language learner. Then reanalysis occurs and renders the initial structures more transparent. Reanalysis of this kind often causes, among other things, category change. For example, the development of the English *for-to* infinitive can be shown to follow from the Transparency Principle (Lightfoot 1979:188):

in the early stage *for* occurs sometimes as a preposition (*I bought it for Mary*, etc.), sometimes as a COMP [(*for to go is necessary)*]; thus the *for* in the new construction of [*for us to go is necessary*] was of unclear category membership and could be analysed as a preposition or as a COMP, causing potential confusion for the language learner. This indeterminancy was removed by the final change, whereby all *for*s were levelled and analysed as prepositions.

I think this Transparency Principle is also relevant in the present discussion about the development of Mandarin *shi*. As will be shown in 2.2.1, at a certain stage in its development, *shi* was analysable either as a demonstrative pronoun or as a copula, causing possible confusion in the language learner. The problem was later on solved by reanalyzing *shi* as a copula.

The Transparency Principle can be shown to be at work in another respect, viz. minimally complex derivation. In modern Mandarin adverbs like *zhen* ('really') and *dou / jie* ('all') obligatorily precede the copula *shi* in equational sentences, but such adverbs follow the pronoun *shi* in Archaic Chinese. To wit

(17) a.

Wang	yue:	'*shi*	*zhen*	wu	shou	fa	zhi	chen	ye.'
king	say	pron	really	my	abide	law	MM	subject	dp

'The king says: "he is really a law-abiding subject of mine." '

(*Han Fei Zi*: 3rd c. B.C.)

b.

qi	qiu	shi	wu	yi.	*Shi*	*jie*	qiu	ming	shi	zhe	ye.
Pron	seek	truth	not	prt	Pron	all	seek	fame	such	person	dp

'They are not seeking truth. They are all after fame as a matter of fact.'

(*Zhuang Zi*: 4th c. B.C.)

The adverbs *zhen* and *jie* follow the pronoun *shi* in (17a) and (17b), respectively. In modern Mandarin, such adverbs will precede the copula *shi*, or the CFM *shi*, as in (18).

(18)

Zhen	*shi / *shi zhen*	ta	shuo	de.
really	CFM	he	say	MM

'It is really he who said it.'

We may use this as a parameter and label it 'pre-ADV'. Recall the pre-NP constraint. If we view it reversely in terms of compatibility of the demonstrative, the copula or the CFM with a following non-NP, what we now have is (19):

(19)

	dem	cpl	CFM
pre-non-NP	–	–	+
pre-ADV[3]	+	–	–

According to (19), the demonstrative *shi* differs from the CFM *shi* by two structural features but differs from the copula *shi* only by one. Since both (16a) and (16c) involve the change dem > CFM, a process that would have involved a more complex derivation than that represented by (16b), which involves only one feature-value change, by the minimal complexity requirement of the Transparency Principle, (16b) would be the most appealing process of diachronical change. The closer diachronical relationship of the demonstrative to the copula than to the CFM, as suggested by (16b), is borne out by the fact that the

structural similarity between the demonstrative and the copula outweighs that between the demonstrative and the CFM, as shown above.

2.2 DIACHRONICAL EVIDENCE

So far, we have mainly examined synchronic facts in arguing for the process dem > copula > CFM. In doing so we have assumed that the demonstrative is older than the other two and suggested that the copula *shi* derived from the demonstrative *shi* historically. Diachronical evidence for this claim is as follows.

2.2.1 In Archaic Chinese, equational sentences normally did not have a copula (Wang 1984), as for example:

(20) a. qi mu yue: 'Kongzi xian ren ye.'
his mother say Confucius wise man dp

'His mother says: "Confucius (is) a wise man." '

(*Zhan Guo Ce*, 5th c. B.C.)

b. wu ba zhe, san wang zhi zui ren ye.
five hegemon person three king Gen criminal person dp

'The five hegemons are criminals before the three kings.'

(*Mencius: Gao Zi Xia*, 4th c. B.C.)

The modern Mandarin copula, *shi*, however, occurs regularly in equational sentences. This modern copula was a demonstrative in Archaic Chinese (Wang 1984), which could be used as a deictic pronoun (21a), or a resumptive pronoun (21b)[4]:

(21) a. fan *shi* bu si, yi yi yan zai!
contrary this not think just it prt Int.

'Do not think contrary to this, that's it!'

(*Shi: Wei Feng: Mang*, 12-6th c. B.C.)

b. zhi zhi wei zhi zhi, bu zhi wei bu zhi, *shi* zhi ye.
know it as know it not know as not know this know dp

'To show that you know when you know and to show that you don't know when you don't, this is true knowledge.' (*Analect*, 5th c. B.C.)

Sentences like (21b) can be interpreted as having a topic-comment structure like (22):

(22) [S1] [*shi* S2]

dem

TOPIC COMMENT

where *shi* as a resumptive pronoun is the subject within the comment part. The reason for such an analysis is that in the period of Archaic Chinese, equational sentences as a rule did not use a copula, as shown in (20), and that there is no sentence where *shi* functions unequivocally as a copula verb.

Then, at a later stage, the topic-comment construction without a copula was *reinterpreted* as a subject-predicate construction with the resumptive pronoun now being reanalysed as a copula (L & T 1975:424). This process of grammaticalization of the topic can be shown as (23):

(23) TOPIC - COMMENT --> SUBJECT - PREDICATE

[S1] [*shi* S2] [S1] [*shi* S2]

dem --> cpl

This process of reanalysis first occurred with the resumptive pronoun *shi*, and subsequently probably through grammaticalization extended the copula use to the deictic *shi*. By the late Han period(1st - 2nd c. A.D.), the use of *shi* as a copula was firmly established and became productive:[5]

(24) a. 'ci bai wu shi he deng?' da yue: 'ci *shi* xiao er.'
this white thing cpl what kind answer say this cpl small child

' "What kind of thing is this white stuff?" Answer: "it's small child." '

(Translation of the Buddha's saying in late Han, 25-220 A.D.)

b. yu *shi* suo jia furen zhi fu ye.
I cpl pron marry woman Gen father dp

'I am the father of the woman given in marriage.'

(*Lun Heng* by Wang Chong, 1st c. A.D.)

In all the sentences above, *shi* is unquestionably copular, in that it is impossible to interpret *shi* as a demonstrative pronoun. For example, in (24a), if *shi* were a demonstrative, we would have two adjacent demonstratives, *ci* and *shi*, which is quite redundant and implausible. The copular function of *shi* in (24b) is obvious for similar reasons, and especially so when we compare (24b), which requires a grammaticalized copula, with an almost identical sentence (24c) from an older text, where the copula does not appear:

(24) c. yu er suo jia furen zhi fu ye (from L & T 1975)
I you pron marry woman Gen father dp

'I am the father of the woman you gave in marriage.'

(*Zhuo Zhuan*, 6-7th c. B.C.)

To sum up, the diachronical development of *shi* in Chinese shows a process of reanalysis whereby an original resumptive pronoun in the comment section of a topic-comment construction was reinterpreted as a copula in the predicate of a subject-predicate construction. Thus, the *demonstrative > copula* part of the hypothesized change is verified.

2.2.3 Finally, let's turn to the evidence in support of the suggested derivational process whereby the CFM *shi* derived from the copula *shi*. It is commonly believed that *shi* as a CFM made its first appearance in the popular verses of the Yuan Dynasty (1271-1368 A.D.), and by the Ming Dynasty(1368-1644) such uses had become very productive. In my research, however, the use of *shi* as a CFM may be dated as far back as the Tang

Dynasty (618-907 A.D.), when examples of the following kind could be found in the poetry then:

(25)

ming	hui	bu	gan	bian,	xin	zhuan	shi	*shi*	nan
fame	return	not	dare	argue	heart	turn	really	CFM	difficult

'(I) do not dare to argue for my reputation, but it is indeed difficult to change my mind.' (Meng Jiao, *Xi Ku*)

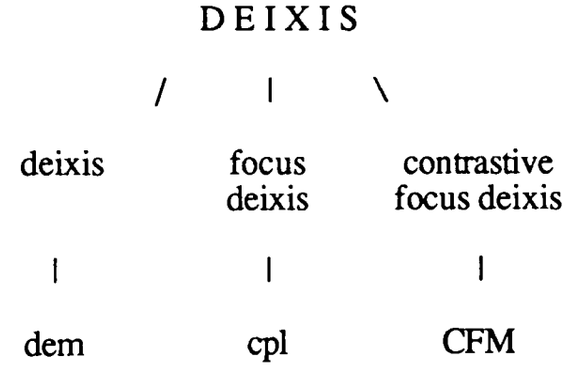
where *shi* appears before an adjective and functions as a CFM, contrastive in the same sense as we discussed with regard to (12).

3. COMMON PRAGMATIC FUNCTION

On the basis of both synchronic and diachronic evidence examined, it can be concluded that in Mandarin there has been historical linkage between the (now obsolete) demonstrative, the copula, and the CFM. As exemplified at the beginning of this paper, there are other languages where the CFM takes the same form as the copula and/or the demonstrative. This phenemenon begs the question from a functional point of view: is this recurrent identity of form between different grammatical categories a mere coincidence, or is there any common functional ground underlying such identity?

Proceeding from the assumption that 'the coding relation between structure and function in syntax is *non-arbitrary*' (Givón 1984:33), and that 'recurrent identity of form between different grammatical categories will always reflect some perceived similarity in communicative function' (Haiman 1985:19), I propose (26) as a common communicative function underlying the demonstrative, the copula, and the CFM, which motivates the formal identity in question.

(26) The demonstrative, the copula, and the contrastive focus marker all share the functional domain of *deixis*, differing from each other only in their respective sub-domains, which can be represented as



3.1 That demonstratives, especially deictics, have a deixis function is self-evident: prototypically, by pointing to some person or entity, the speaker clearly identifies the referent in the immediate environment. However, since what is being deictically identified can occur either in a focal position (e.g. object) or in a non-focal position (e.g. subject), we may neutralize the distinction between focus and non-focus deixis by just saying that demonstratives usually serve to fulfil the function of deixis. It should be noted, however, that this is so only in unmarked cases where the demonstrative is unstressed; when a demonstrative is stressed, however, it can assume the function of contrastive focus deixis.

3.2 If we define focus at least partially as that part of sentence which carries new information, the copula may then be said to signal that what follows it (or precedes it, as in Japanese) is the focus/new information, and thus helps to deictically identify focus. As for the identification function, we have to distinguish between two kinds of predicates that can follow the copula. Consider the following:

(27) a. That man is (a) nice (man).

b. That man is the suspect.

In (27a) the NP/AP following the copula characterizes the subject NP as possessing a certain property, and no identification is involved. We may call such predicates 'predicational'. In (27b), however, the referents of two definite terms coincide in the same person/entity, and the relationship between the two NPs is one of identification. We may call such predicates 'identificational'. In such sentences, the subject NP may be called 'the identified', the post-copula NP 'the identifier' (Halliday 1976), and the copula 'identifying be', [which] 'describes a State in which a relation of identity is established between two entities described by means of two different terms' (Dik 1985:32). In this sense, the copula serves to deictically mark the identifier, which is the focus of the sentence.

3.3 While the copula has a function of focus deixis, the term CFM itself suggests that it has a function of *contrastive* focus deixis. In an English sentence like (1), the contrastive focus can be said to be marked both by the structural configuration and by the copula-like CFM. In languages where the focused constituent is in the same position as in a corresponding non-cleft sentence, such as Mandarin and Mopun, the CFM is the sole means available to mark the contrastive focus and has a more crucial role to play in deictically identifying the contrastive focus. However, since identification normally relates two *definite* entities, what is being focused upon usually 'is not new in the sense that it is completely out of consciousness. What is new is that [the element being focused] is the particular one which fits the description in [the clause which contains backgrounded information,] i.e. it is chosen as the correct object from a closed set of given possibilities and is therefore proclaimed as the correct element of such an exhaustive listing' (Brömser 1984:329; Chafe 1974). Thus cleft sentences belong to the 'special and readily identifiable class'(Chafe 1974:118) 'which express[es] a *contrast*' (Brömser 1984:329) (italics mine). The CFM in such sentences helps to single out the element in contrastive focus in an unambiguous way.

4. CROSSLINGUISTIC EVIDENCE

Having examined Mandarin in some detail in proposing a functional account for the formal identity among the demonstrative, the copula and the CFM, we will now turn to some other languages to see if such cross-categorial formal identity occurs recurrently. Four languages will be discussed: Hebrew, Margi (a Chadic language spoken in West Africa), Mokilese, and Kusaiean, the latter two being both Micronesian languages.

4.1 HEBREW

In present tense equational sentences in Hebrew, a demonstrative *ze* or a personal pronoun *hu* is usually used to replace the otherwise used copula *h_y_y* (Berman 1978; L & T 1975):

(28) a. she-hu axrai lamaase (ze) barur lexulam
that-he responsible for-the-dead dem clear to-all
'That he is responsible for the dead is clear to all.'

b. ata (hu) ha-ganav
you he the-thief
'You are the thief.'

L & T (1975) parallels *hu* to the Mandarin *shi* by suggesting that it is being reanalyzed as a copula in modern Hebrew. This is seen in several structural aspects:

1) Sentences like (28) can be uttered with an intonation suitable for a single, simple sentence, without a break after the subject, a feature that characterizes a subject-predicate structure.

2) The subject of a subject-predicate equational sentence can be indefinite, but the topic in a topic-comment sentence cannot.

3) *hu* can appear with a non-third person subject (e.g. (28b)), in which case *hu* cannot possibly have a 3sg referent and has to be analysed as a copula.

As for *ze*, I will present the following argument in terms of agreement, to suggest its copular status. In declarative equational sentences, *ze* agrees in gender with the predicate NP if the latter is predicational. Consider (29):

(29) ha-bayit shelha *zot* *dugma* *tova* [casual] (Glinert 1989:189)
the-house(m) your cpl(f) example good(f)
'Your house is a good example.'

zot in (29) is the feminine inflection of *ze*. It agrees in gender with the feminine predicational NP rather than with the subject NP. Given the fact that deictic adjectives generally follow the head noun in Hebrew, if *zot* were a true demonstrative, the head noun it modifies would be the subject NP, and we would expect it to agree with the masculine subject NP and take the form *ze*. That this is in fact not the case suggests that *ze*, like *hu*, has been reanalysed as a copula.

While L & T (1975) only mentions the optionality of the pronouns in sentences like (28a) and (28b), Berman (1978:207-8) points out that in a sentence like (28b), if *hu* does appear, 'the effect achieved by having a 3rd-person pronoun 'copy' [=copula in L & T] following a non-3rd-person Subject is of a kind of 'clefting', with its concomitant emphasis manifested by the obligatory contrastive stress on the Subject pronoun.' For example,

(30) a. **ata** hu hexashud
you cpl the-suspect
'*You* are the suspect / It is you who are the suspect.'

b. **ani** hu she- pihakti (Glinert 1989)
I cpl who I-yawned
'It is I who yawned.'

The structural clefting with *hu* as the CFM is more clearly shown in (30b) , which, by using a relative pronoun *she-*, shows a bi-clausal structure.

Like *hu*, *ze* also can be used as a CFM in cleft sentences to mark contrastive focus, and when thus used it precedes the focused constituent (Glinert:422f):

(31) a. *ze* bney-adam tovim o bney-adam ra'im she-rotsim et ze?
CFM people good or people bad who-want OM it
'Is it good or bad people that want it?'

b. *ze* oti ra'it ba-shuk [infml]
CFM me you-saw in-the-market
'It is me you saw in the market.'

c. *ze* ani she-pihakti
CFM I who-yawned
'It is I who yawned.'

(31a) and (31b) show, respectively, that the focused constituent can be either the subject or the object in the clause. (31a) and (31c) show that the relative marker *she-* can head the presupposed clause.

In sum, Hebrew parallels Mandarin in terms of both cross-categorial formal identity and diachronical relationship between the categories, though the exact path of derivation involving the CFM is less clear. The difference seems to be that while Mandarin subsequently lost the pronominal function of *shi* following its reinterpretation as a copula, Hebrew nevertheless has retained the pronominal function for *ze / hu*.[6]

4.2 MARGI

In Margi, the common demonstrative pronouns are *ku* 'this', *ta* 'that', etc. There also exists a morpheme *nu*, whose deictic use is now found only in a few words, for example, *uven* 'here' (probably from *u* 'in' + *ivi* 'place' + *nu* 'this'). Apart from its deictic meaning, its structural properties also show that this morpheme was originally a demonstrative. Structurally, *nu* parallels *ku* in that a) both can be used to nominalize adjectives, as in (32), and b) both can be used as a relative pronoun if the relativized NP is the subject in the relative clause, as in (33).

(32) kə/nə dəgal 'the big one', and

(33) a. naja ga kətər mdə *ku* at dəar dəl
he saved man who fell-into-river
'He saved the man who fell into the river.'

b. ... ular ja mji mətlu *nu* wudə mama
... saw two people who brothers
'He saw two people who were brothers.'

Margi normally does not use a copula in equational sentences, the predicate NP usually following the subject directly. But sometimes, either the 3sg. pronoun *naja* or the

demonstrative *nu* may occur between subject and predicate 'as a quasi-copula' (Hoffmann 1963:275):

(34) a. hya ku *naja* mala
dog this cpl bitch
'This dog is a bitch.'

b. kəkə'yar kə *nu* mənagu
these cpl good
'These are good.'

c. ja-*n*
it cpl
'That's it!'

In (34c), where no expletive subject NP is required, only the predicate NP is obligatory, and the copula *nu* occurs at the end.

Like *shi* in Mandarin and *ze* in modern Hebrew, the demonstrative pronoun *nu* in Margi has probably been reanalysed as a copula. There are at least three pieces of evidence for this. First, with nominal predicates, *nu* can follow plural NPs, as in

(35) nanda nu 'It is they.'

Given the existence of plural demonstratives *kuku'yr* or *kuku'yar ku* 'these', and given that a demonstrative agrees in number with the noun it modifies, if what follows the NP in (35) were a true demonstrative, one of the plural demonstratives should have appeared instead of the singular *nu*. The fact that what actually shows up is the latter suggests that this morpheme has probably lost most of its demonstrative function and been reanalyzed as a copula. Besides, it is rather unusual for a personal pronoun, which already has definite and complete reference, to be further designated by a demonstrative. Secondly, (35) would not be a sentence if *nu* were not a copula-like verb. Thirdly, *nu* is not mutually exclusive with the more common demonstrative *ku*. Consider (34b) again. Given that in general members of the same grammatical paradigm tend to be mutually exclusive with regard to a certain syntactic position, if *nu* were a genuine demonstrative, we would have two adjacent demonstratives with the same meaning and function, which is both structurally and pragmatically not very plausible, not to mention their disagreement with each other in number. On the other hand, if we analyse *nu* as a copula, there will be no such problems.

Hoffmann (1963:89) also notes that, from the use of *nu* (as a copula) in nominal sentences is 'derived the use of *nu* as a particle, introduced between subject and predicate, in order to give the subject a certain emphasis.' In fact, a cleft reading is possible:

(36) a. ni nu samən
I CFM brother
'*I* am your brother / It is I who am your brother.'

b. ni -n atsianyi
I CFM kill-him
'*I* killed him / It is I who killed him.'

Such emphatic use of *nu* corresponds to the CFM use of *ze/hu* in Hebrew (cf. (30-31)). In addition, the whole nominal predication in (34c), *ja-n*, in which *nu* is used like a copula, may also be used as a CFM. For instance,

(37) kər banban jan attəny
headache CFM hurt-him
'*Headache* hurts him / It is headache that hurts him.'

(37) provides us with a clue in favour of the possible diachronical change *dem* > *copula* > *CFM*. Making the reasonable assumption based on the non-use of a copula in general that the now obsolete demonstrative use precedes the copula and CFM uses historically, if the ordering were *dem* > *CFM* > *copula*, then we would expect the CFM form in (37) to be *nu*, a direct derivation from the demonstrative form, rather than *jan*, a form in which *nu* is used as a copula. The fact to the contrary suggests that CFM use must have derived from the copula use, at which stage *nu* is suffixed to *ja*, and that only at a later stage did the whole fused form begin to be used as CFM. Therefore, the diachronical change is more likely to be *dem* > *copula* > *CFM* in Margi, just as in Mandarin.

4.3 MOKILESE

Mokilese has a basic SVO order, and any major constituent, including the VP[7], can be cleft focused by being placed at the beginning of the sentence, with the remainder of the sentence treated as a relative clause and introduced by the relative marker *ma* (Harrison 1976):

(38) a. ngoah wahdo mwingehu
I bought food
'I bought the food.'

b. ngoah [ma wahdo mwingehu]
I RM bought food
'It was I who bought the food.'

Parallel to this cleft construction is an alternative cleft pattern (39), where the initial cleft focus is preceded by *ioar* which functions as a CFM:

(39) a. ioar woall-o ma wia mehu
CFM man-that RM did that
'It was that man who did it.'

b. ioar kida ma ngoah nimen dupukda
CFM guitar RM I want buy
'It's a guitar that I want to buy.'

According to Harrison (1976:309), *ioar* is probably 'the last survivor of a set of 'pointing determiners,'' i.e. deictic pronouns which have become obsolete[8]. In fact, even this last survivor is now being used so rarely that evidence of its synchronic demonstrative use is lacking from the available Mokilese data per se. However, its cognates *iet*, *ien*, and *io* in Ponapean, another language in the same Ponapeic subgroup of the Micronesian languages, are still being used as demonstratives (*ibid.*), and are glossed as such in the list of cognates between Kusaiean and Ponapean provided by Lee (1975), where *io* (spelled as

o) and *ien* are presented as cognates to the Kusaiean demonstratives *ah* and *an*, respectively. Thus, the Mokilese, Ponapean and Kusaiean cognates for the demonstrative can be shown as follows:

(40)

Dem	**Mokilese**	**Ponapean**	**Kusaiean**
'this'	ioar	io	ah
'that'	?	ien	an

The morphophonemic alikeness between the Mokilese *ioar* and the Ponapean *io* would become even more transparent if we, by assuming a bimorphemic structure for *ioar*, could analyse it as composed of *io* 'that', which is morphophonemically identical with its Ponapean cognate, and some other morpheme. This analysis is supported by the existence of a delimiting suffix *-oar* in Mokilese, which is translated as 'only' and suffixed directly to a NP or a NP modifier such as a demonstrative. It is this delimiting suffix, I suggest, that constitutes the latter part of *ioar*, i.e. *ioar* may well be analysed as *io-oar* 'this-only'. Given its use as a CFM, which identifies as contrastive focus a constituent in a sentence whose referent tends to have exclusive reference in the given discourse, this analysis is semantically quite plausible[10]. In short, the Mokilese CFM *ioar* can be analysed as involving a demonstrative whose use has become largely obsolete.

In addition to its (obsolete) demonstrative use, it is also found that sometimes, *ioar* can be used as a copula in Mokilese equational sentences with definite predicate NPs. In general, an equational sentence may be either predicational or identificational, depending on whether the predicate NP is indefinite or definite. Mokilese identificational equational sentences allow for a copula. Consider

(41) a. John johnpadahkmen
John teacher
'John is a teacher.'

b. minpas-e ioar noai pinjel-wa
one-this cpl my pencil-Det
'This one is my pencil.'

c. me ioar woal-wa ma wia mehu
this cpl man-the RM did that
'This is the man who did it.'

The predicate NP in (41a) is predicational by virtue of being indefinite, and *ioar* is not used. The predicate NPs in (41b-c), on the other hand, are both definite referential NPs; and between the subject and the predicate appears *ioar*, whose function is clearly that of a copula. Furthermore, its copular status is vindicated by the same argument as presented in the Margi case with regard to the property of mutual exclusiveness between members of the same grammatical paradigm. More specifically, if *ioar* in (41b) or (41c) were analysed as a demonstrative, there would coexist two adjacent demonstratives for no obvious reasons at all, which renders the analysis implausible. The copula analysis, as in the Margi case, eliminates such problems. Thus, a relationship of polysemy obtains whereby the same form, *ioar*, can function as a CFM, a demonstrative, or a copula depending on the syntactic position in which it occurs in a sentence.

There is no direct evidence for diachronical relationships between the three categories in Mokilese. However, I will venture the following speculations. First, if we agree that where polysemy exists between a form and two or more related functions the obsolete function tends to diachronically precede the synchronically active ones, then the obsoleteness of *ioar* as a demonstrative may suggest its diachronical precedence to the other two functions. Secondly, we notice in (39) that when *ioar* is used as a CFM, it occurs to the left of the focused constituent, contrary to our usual expectation of a VO language. On the other hand, the demonstrative *ioar*, by analogy to other currently operative demonstratives, and the copula *ioar* occur in positions predictable for a VO language. Probably, this reflects an on-going word order change in Mokilese. What this implies is that the CFM function of *ioar* is probably a more recent innovation than the copula function. Therefore, the diachronical development of the functions of *ioar* can be tentatively shown as *dem > copula > CFM.*

To summarize, despite the speculative nature of the argument with regard to the diachrony, one thing that is undoubtedly conclusive is the close relationship between the three functions in question, all of which are realized by the same form.

4.4 KUSAIEAN

Kusaiean, another Micronesian language with a basic SVO order, presents a somewhat different case. In Kusaiean, the cleft focus occurs in sentence-initial position, followed immediately by *pa*; and the whole sentence must end with the particle *ah* or *uh* (Lee 1975):

(42)

wes	se	**pa**	nga	enenuh	**uh**
shirt	this	cpl	I	need	Prt

'It is the shirt that I need.'

Two things are worth noting in (42). First, *pa* is glossed as 'copula', as can be seen from its occurrence in identificational equational sentences. As in Mokilese and some other languages, Kusaiean identificational equational sentences differ structurally from their predicational counterparts in having a copula, *pa*, between the subject NP and the predicate NP:

(43) a.

nga	tuhlihk lutlut	se
I	student	a

'I am a student.'

b.

mwet	sac	*pa*	mwet pihsrapasr	sac
man	that	cpl	thief	that

'That man is the thief.'

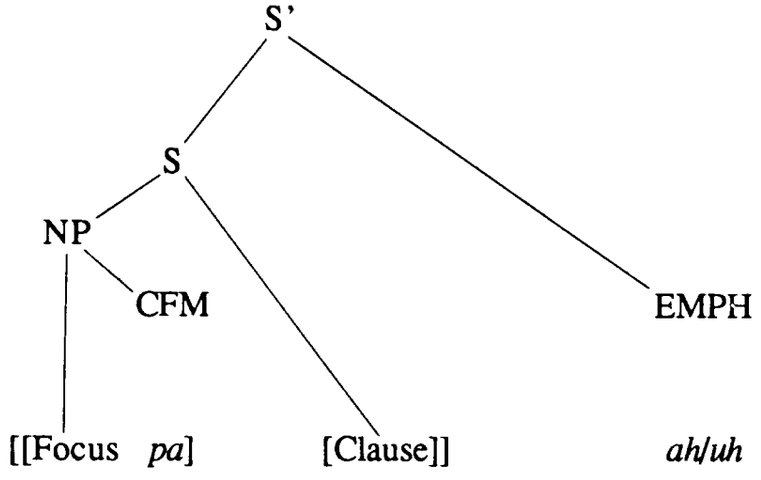
The predicate NP in (43a) is predicational by virtue of being indefinite, and no copula is used. *Pa* is present in (43b) because the predicate NP is definite and therefore identificational. The function of *pa* in (43b) is clearly copular. This same form is used in cleft sentences where it occurs adjacent to the cleft focus and poses as a candidate for the CFM[11].

Secondly, the sentence-final particle *ah* or *uh*[12] occurs *obligatorily* in sentences like (42), and poses as another potential CFM. This *ah/uh* is in fact a demonstrative, which use

is self-evident from such examples as *ik uh* 'these fish' and *infacl ah* 'in the river'. As for the relationship between *pa* and *ah/uh*, there is no evidence suggesting their diachronical relatedness. Therefore, the two grammatical categories, copula and demonstrative, are realized by two different morphemes in Kusaiean.

Given the situation that a Kusaiean cleft sentence requires both a copula-like form and a demonstrative pronoun and either can be a candidate for the CFM, which of the two actually serves as the CFM? Considering the close linguistic distance between the focus marker and the focused constituent observed in many languages, it would seem more likely that the copula serves the function of the CFM, which is in part supported by Lee's (1976) glossing of *pa* as 'focus marker'. However, the obligatory presence of the demonstrative in cleft sentences still remains to be accounted for. I suggest that both serve to mark contrastive focus but at different levels, i.e. the sentence-final demonstrative/particle marks, on a more general level, the whole sentence as contrastively focusing on one of its constituents, whereas the copula marks a specific constituent as the contrastive focus per se. Otherwise, it would seem rather implausible for *ah/uh*, obligatory as it is, to mark a distant focus. If this analysis is adopted, we would have a cleft structure like (44):

(44)



in which the copula and the demonstrative pronoun, though diachronically unrelated, jointly mark a cleft sentence but each operates at a different level.

5. CONCLUSION

It has been shown that formal identity obtains recurrently between the three grammatical categories crosslinguistically, as in (45).

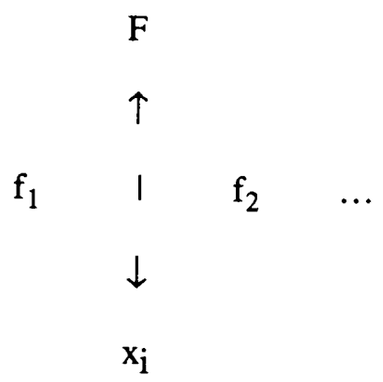
(45)

	Mandarin	**Hebrew**	**Margi**	**Mokilese**	**Kusaiean**
CFM	shi	ze/(hu)	nu	ioar	pa...ah/uh
cpl	shi	ze/(hu)	nu	ioar	pa
dem	shi	ze	nu	io	ah/uh

Now turning back to (9), we have seen that both diachronical relatedness (9b) and functional commonality (9c) are responsible for the recurrent formal identity between the different grammatical categories. On the one hand, these languages, again with the exception of Kusaiean, manifest to different extents diachronical relationships between the

three grammatical categories. On the other hand, we can perceive a pragmatic function of *deixis* underlying all these categories in all the languages, each with its own sub-functional domain. It is this underlying function, I think, that historically motivated the formal identity between the copula, the CFM and the demonstrative pronoun in languages that exhibit such identity. The function proposed is a kind of arch-function in the sense that it is underspecified for unifying several sub-functions, each represented by a different grammatical category. The whole matter can be perceived as shifting sub-functional domains within a general functional domain while keeping the form intact. We can envisage a function hierarchy whose relation with the same recurrent form can be diagrammed as (46):

(46)



where F = arch-function, f = interrelated, specific function, and x_i = recurrent identical form. The relationship of x_i to f_n is one of polysemy, i.e. related functions mapped onto the same form. The relationship of x_i to F at a higher level is one of isomorphism. Therefore, the whole issue seems to boil down to the underlying isomorphemic principle of one form for one function/meaning, which, by manifesting itself at a more abstract level, motivates the phenomenon in question.

The conclusion drawn from this study has a certain measure of crosslinguistic validity, which may inspire future studies to see if the same cross-categorial formal identity exists in other languages and if the results conform to the claim made herein. One such study that touches upon the issue is already available in Wolvengrey (1990, ms.) on an Amerindian language Mandan, which presents some evidence of a general demonstrative *-e*, augmented by *-n*, that has been extended from its original function as a demonstrative to that of a grammaticalized CFM. It is hoped that the present study will be helpful to such future studies on similar issues.

ACKNOWLEDGEMENTS

I wish to thank John Haiman for his advice and many invaluable comments and suggestions on earlier drafts. I am grateful to Richard Carter for many useful discussions, and T. Russel for consultations on Archaic Chinese data. I am also indebted to Moshe Nahir for his useful discussions on the Hebrew data. I am of course solely responsible for the final interpretations and remaining errors.

FOOTNOTES

1 In a traditional transformational approach, derivation of a cleft sentence as here presented would have to involve movement by virtue of left dislocation (Akmajian 1970). In other more recent generative approaches, such as GPSG, clefts of this kind would be base-generated by virtue of a

PS-rule like *S -> cpl F S'* (Knowles 1986). Without involving myself in the dispute which is irrelevant to the issue expounded in this paper, I simply describe an obvious structural property of the clefts here.

2 Even if without *shi*, **ta cong Zhongguo* is still ungrammatical without another verb that follows it. This, however, does not invalidate our argument here, since a sentence like ?**ta shi cong Zhongguo lai*, in which *shi* is used as a 'pure' *copula*, is still problematic.

3 'Pre-NP' here only refers to the intensive adverbs such as those mentioned above.

4 Even today, remnants of *shi* as deictics can still be found in pedantically formal writings or idioms, for example,

(i)
shi	ri
that	day

'that day'

(ii)
shi	ke	ren,	shu	bu	ke	ren?
this	may	tolerate	what	not	may	tolerate

'(If) this is tolerable, what cannot be?'

5 The exact time for the demonstrative -> copula change is not clear. In my research, however, sporadic examples indicating the beginning of such change are found in the works of as early as the third century B.C., a time when equational sentences regularly did not require a copula. Minimal-pair examples like

ci	he	zhong	ye ...	ci	*shi*	he	zhong	ye.
this	what	kind	dp	this	cpl	what	kind	dp

'What kind (is) this? ... What kind is this?'

(*Han Fei Zi*: 3rd c. B.C.)

clearly suggest the copular status of *shi*. However, since examples like this are very rare and sporadic, they do not count as reflecting a regular copula function of *shi* in the works of that period.

6 That *ze* and *hu* still function as pronouns is seen in examples like *ze meshune* 'that's strange'. Also worthy noting here is Doron's (1986) argument that the pronoun form *hu* in Hebrew is not the present tense suppletive form of a verbal copula which is part of the predicate, but a clitic in INFL, i.e. 'the phonological realization of 'unattached' agreement features that have absorbed Case' (p.313) (cf. L & T (1975) on Hebrew and Eid (1983) on Arabic.

7 When the main verb of the sentence, or some structure containing the main verb, is cleft focused, the main verb position must be filled by the pro-verb *wia*:

loakjid	ma	arai	pirin	wia	rehnnoawe
to-fish	RM	3pl	Aux	do	today

'It is fishing that they're going to do today.'

A similar kind of verb clefting can also be found in Breton (Timm 1987).

8 The currently productive demonstratives in modern Mokilese include, among other things, *-e/-i* 'this', *-en/-n* 'that' and *-o/-u* 'that (remote)', which are used as suffixes:

woall-e	'this man'
woall-en	'that man (near you)'
woall-o	'that man (over there)'

9 Not surprisingly by now, the Kusaiean demonstrative *ah* can also be used as a CFM, as will be discussed shortly.

10 In this respect, *ioar* may have been used as a sort of emphatic determiner, like the English *the* in 'He is *the* man.'

11 However, *pa* does not always occur in cleft sentences: if the cleft focus is a question word corresponding to an English WH-word, *pa* simply does not appear. In fact, this is true of all interrogative sentences with a question word:

a. suc (*pa) tuhkuh ah?
who cpl came prt

'Who is it that came?'

b. suc (*pa) el an? - el pa Sah
who he prt he cpl Sah

'Who is he?' - 'He is Sah.'

12 The only difference between *ah* and *uh* seems to be temporal: the former cooccurs with a past tense verb in the presupposed clause, while the latter with a non-past verb

ABBREVIATIONS

Asp	aspect
cpl	copula
CFM	contrastive focus marker
cl	classifier
dem	demonstrative
dp	declarative particle
f	feminine
Gen	genitive
Hon	honorific
Loc	locative
m	masculine
M	measure word
MM	modifier marker
OM	object marker
Perf	perfective aspect marker
pl	plural
pron	pronoun
prt	particle
RM	relative marker
S	sentence
Sfx	suffix
TOP	topic

REFERENCES

AKMAJIAN, A. 1970. 'On deriving cleft-sentences from pseudo-cleft sentences.' *Linguistic Inquiry* 1: 149-68.

BENNETT, P.A.(1979). 'Observations on the transparency principle.' *Linguistics 17*: 301-15.

BERMAN, R.A. 1978. *Modern Hebrew Structure*. Tel-Aviv: University Publishing Projects Ltd.

BRÖMSER, B. 1984. 'Towards a functional description of cleft constructions.' *Lingua 62*: 325-48.

CHAFE, W.L. 1974. 'Language and consciousness.' *Language* 50, 111-33.

--------. 1976. 'Givenness, contrastiveness, definiteness, subjects, topics, and point of view.' In C. Li (ed.), *Subject and Topic*, 25-56. New York: Academic Press.

CHAO, Y-R. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

CHEN, J. 1984. *Hanyu Kouyu* (Spoken Chinese). Beijing: Beijing Press.

CHOMSKY, N. 1957. *Syntactic Structures*. The Hague: Mouton.

CRAIG, C.G. 1977. *The Structure of Jacaltec*. Austin: University of Texas Press.

DIK, S.C. 1978. *Functional Grammar*. Amsterdam: North-Holland Publishing Co.

-------- 1985. 'Cleft and pseudo-cleft in functional grammar.' In W. Zonneveld & F. Weerman (eds.), *Linguistics in the Netherlands 1977-1979*, 26-43. Dordrecht: Foris.

DORON, E. 1986. 'The pronominal 'copula' as agreement clitic.' In H. Borer (ed.), *Syntax and Semantics*, Vol.19: 313-32. Orlando: Academic Press.

EID, M. 1983. 'The copula function of pronouns.' *Lingua 59*(1): 197-207.

ELLIS, J. & BOADI, L. 1969. '*To be* in Twi.' In J.W.M. Verhaar (ed.), *The Verb 'Be' and Its Synonyms*, Vol. 9, 1-71. Dordrecht: D. Deidel Publishing Co.

FRAJZYNGIER, Z. 1987. 'From preposition to copula.' in D. Odden (ed.), *Current Approaches to African Linguistics* IV: 371-86. Dordrecht: Foris.

GILI, J. 1974. *Introductory Catalan Grammar*. Oxford: The Dolphin Book Co..

GIVON, T. 1979. *On Understanding Grammar*. New York: Academic Press.

-------- 1984. *Syntax: A Functional-Typological Introduction*. Amsterdam: Benjamins Publishing Co.

GLINERT, L. 1989. *The Grammar of Modern Hebrew*. Cambridge: Cambridge University Press.

GREENBERG, J.H. 1966. 'Some universals of grammar with particular reference to the order of meaningful elements.' In J.H. Greeneberg (ed.), *Universals of Language*, 58-90. Cambridge, Mass.: MIT Press.

--------. 1985. 'Some iconic relationships among place, time and discourse deixis.' In J. Haiman (ed.), *Iconicity in Syntax*, 271-87. Amsterdam: John Benjamins Publishing Co.

HAIMAN, J. 1974. *Targets and Syntactic Change*. The Hague: Mouton.

--------. 1985. *Natural Syntax*. Cambridge: Cambridge University Press.

HALLIDAY, M.A.K. 1976. 'Notes on transitivity and theme in English.' Part 2. *Journal of Linguistics* 3: 199-244.

HARRIES-DELISLE, H. 1978. 'Contrastive emphasis and cleft sentences.' In J.H. Greenburg (ed.), *Universals of Human Languages* Vol. 4, 419-86. Standford, California: Standford University Press.

HARRISON, S.P. 1976. *Mokilese Reference Grammar*. Honolulu: The University Press of Hawaii.

HASHIMOTO, A.Y. 1969. 'The verb 'to be' in modern Chinese.' In J.W.M. Verhaar (ed.), *The Verb 'Be' and Its Synonyms*, Vol. 9, 72-111. Dordrecht: D. Reidel Publishing Co.

HAWKINS, J.A. 1988. *Explanation of Universals*. Oxford: Basil Blackwell.

HEGGIE, L.A. 1988. *The Syntax of Copular Structures*. Doctoral dissertation. Los Angeles: University of Southern California.

HOFFMANN, C. 1963. *A Grammar of the Margi Language*. London: Oxford University Press.

HORN, L.R. 1981. 'Exhaustiveness and the semantics of clefts.' In V. Burk & J. Pustejovsky (eds.), *Proceedings of the 11th Annual Meeting of the North Eastern Linguistic Society*, 125-42. Amherst: University of Massachusets.

HUANG, C.-T. J. 1982. *Logical Relations in Chinese and the Theory of Grammar*. MIT dissertation.

KEENAN, E. & COMRIE, B. 1977. 'Noun phrase accessibility and universal grammar.' *Linguistic Inquiry* 8(1): 63-99.

KNOWLES, J. 1986. 'The cleft sentence: a base-generated perspective.' *Lingua* 69(4): 295-317.

KOOPMAN, W. et al.(eds.) 1987. *Explanation and Linguistic Change*. Current Issues in Linguistic Theory, Vol. 45. Amsterdam: John Benjamins Publishing Co.

LEE, K-D. 1975. *Kusaiean Reference Grammar*. Honolulu: The University Press of Hawaii.

--------. 1976. *Kusaiean-English Dictionary*. Honolulu: The University Press of Hawaii.

LEE, M. 1988. 'Language, perception and the world.' In J.A. Hawkins (ed.), *Explanation of Universals*, 211-46. Oxford: Basil Blackwell.

LI, C. & THOMPSON, S.A. 1975. 'A mechanism for the development of copula morphemes.' In C. Li (ed.), *Word Order and Word Order Change*, 419-44. Austin: University of Texas Press.

-------- 1976. *Subject and Topic*. New York: Academic Press.

-------- 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

LIGHTFOOT, D.W. 1979. *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.

LUO, C. 1988. 'Cleft sentences and word order in Chinese.' ms.

McCLOSKEY, J. 1979. *Transformational Syntax and Model Theoretical Semantics*. Dordrecht: D. Reidel Publishing Co.

ROSS, C. 1983. 'On the function of Mandarin DE.' *Journal of Chinese Linguistics* 11(2): 214-46.

SPAULDING, R.K. 1959. *Syntax of the Spanish Verb*. New York: Holt, Rinehart & Winston.

STEIN, D. 1988. 'Semantic similarity between categories as a vehicle of linguistic change.' *Diachronica* V:1/2: 1-19.

TIMM, L.A. 1987. 'Cleft structures in Breton.' *Word* 38(2): 127-42.

VERHAAR, J.W.M. (ed.) 1969. *The Verb 'Be' and Its Synonyms*, Vol. 9. Dordrecht: D. Reidel Publishing Co.

WANG, L. 1984. *Wang Li Wen Ji* (Works of Wang Li), *Vol.1*. Jinan: Shandong Jiaoyu Chubanshe.

-------- 1985. *Modern Chinese Grammar*. Beijing: Commercial Press.

WOLVENGREY, A. 1990. 'A marker of contrastive focus in Mandan', ms.

ZHOU, ZHONGLING et al. 1982. *Han Fei Zi Suoying* (An Index to Han Fei Zi). Beijing: Zhonghua Shu Ju.

THE SWITCH FROM INFINITIVE TO SUBJUNCTIVE CLAUSES IN ROMANIAN

Virginia Motapanyane
University of New Brunswick / Saint-John

ABSTRACT

This paper focuses on the syntactic properties of subjunctive complement clauses in Romanian. The peculiarity of these structures consists in two empirical facts : (i) the subjunctive cannot alternate with the infinitive in complement position – a common option in other Romance languages; (ii) word order is restricted to VSO in COMP-less subjunctive complements, although the SVO/VSO option is free in all other contexts.

I propose a generative approach to these constructions within the framework of Government and Binding Theory (cf. Chomsky (1981)). The purpose of my analysis is to account for the phenomena in (i) and (ii) through specific constraints applying in the underlying structure of a subjunctive clause. The outcome of this analysis will show that Romanian follows the syntactic pattern of Romance languages in these structural contexts. This conclusion weakens the usual relation between Romanian and Greek, a neighbouring language where a similar switch from infinitive to subjunctive complements took place.

0. INTRODUCTION

In this paper I would like to show that some recent theoretical innovations within the Government and Binding framework permit a principled account for the lack of obviation in Romanian sentences with subjunctive complements, a phenomenon that coincides with the unavailability of infinitive versions in the same context. This issue has been an object of controversy in syntax ever since Farkas's (1982) Ph. D. dissertation.

The investigation carried on here will lead me to revise and tighten current assumptions on X-bar theory, the Null Subject Parameter, Government and Case. I will propose a new analysis of the subjunctive Inflectional Phrase (IP) in Romanian where lack of obviation follows from a peculiar configuration at Deep-Structure (DS). This configuration obtains only when the subjunctive clause occupies a subordinate position.

The major theoretical problem with respect to Romanian subjunctive complements arises in sentences where the subjects corefer :

(1)	pro	Vreau	să	plec	în	vacanţă
	(I)	want	SA	(I) go	on	vacation

I want to go on vacation.

In the same context, Romance languages are 'obviative'; that is, a subjunctive clause is possible only when the subject of the matrix clause and the subject of the complement clause are disjoint in reference, e.g.:

(2) Je veux que Marie parte.
I want that Marie go
I want Mary to go.

If the two subjects are coreferent, as in (1) for Romanian, a 'control' relation between the two subjects is obligatory in Romance (i.e. the complement clause will have an infinitive verb):

(3) a. *Je veux que je parte en vacances.
I want that I go on vacations
I want to go on vacation.

b. Je veux partir en vacances.
I want to go on vacations

c. *pro Deseo que termine esta tesis.[1]
(I) wish that (I)finish this thesis

d. pro Deseo terminar esta tesis.
(I) want to finish this thesis

In the infinitive clauses in (3b, d) the matrix subject controls the empty subject of the embedded clause, analyzed as PRO, with [+pronominal], [+anaphoric] features. The same relation rules out the sentence if the second subject is a pronominal (lexical (3a) or empty (3c)). The ungrammaticality is due to violation of the binding principles.

Contrasting the paradigms in (1) and (3) it appears that subjunctive complements in Romanian must present specific properties that allow for the versions ruled out in French (3a) and Spanish (3c). These properties must also be related to the fact that an infinitival complement as in (3b, d) would not be available in contemporary Romanian.

I will approach the contrast between (1) and (3) on the basis of two assumptions : (i) the loss of infinitive complements led to the use of various types of subjunctive clauses; (ii) the mechanism of subordination has an impact on the internal structure of the subjunctive complement.

These two phenomena can be accounted for if I combine elements of the X-bar theory as presented in Chomsky (1986) with other proposals such as : (i) the theory of *pro* (Rizzi (1986)); (ii) the analysis of IP split into various functional heads (Belletti (1989), Pollock (1989)); (iii) the government relation defined through Relativized Minimality (Rizzi (1989)); (iv) the hypothesis of subjects base-generated inside VP (Kuroda (1986), Koopman (1988), Sportiche (1988)).

The paper is organized as follows : Section I illustrates the distribution of the subjunctive clause in Romanian. I propose a representation of the subjunctive IP in section II. This representation will be judged in contexts of θ-selection in section III; the discussion justifies the irrelevance of obviation to Romanian subjunctive complements. Section IV contains a reference to Greek and Romance languages and concludes the analysis.

I. THE DISTRIBUTION OF SUBJUNCTIVE CLAUSES IN ROMANIAN

The subjunctive has extensively replaced the infinitive in complement clauses beginning with the XVIIIth century. The infinitive was still in use before that period, as illustrated in annals and other documents of the time. e.g.:

(4) Intîmplîndu-se *a venire feciorul nostru,* Lupul Vătaful, pentru trebile noastre... (text XVIIth century, cf. Coteanu (1981)).

happening-it to come our son, Wolf-the Chief-the, for interests-the ours...

So it happened that our son, Wolf the Chief, came in relation with some business affairs...

The infinitive clause italicized in (4) would take obligatorily a subjunctive verb in contemporary Romanian.[2]

The factors that triggered this replacement are still unclear. For historical discussions I refer the reader to Rivero (1987) and Dobrovie-Sorin (1989). The result of this switch consisted in an extended use of subjunctive clauses in various contexts. This situation will be illustrated in the data below.

Before presenting the empirical data, I must draw attention to the fact that word order in these sentences will have a crucial importance for the analysis to be developed in sections II and III. In this sense, I will mention the position of the subject in relation to the verb for every sentence. The subjunctive verb is italicized.

1.1. ROOT CLAUSES

The subjunctive serves as a suppletive form for the imperative in Romanian, as in many other languages. e.g.:

(5) a. *Să-mi aducă* Ion nişte fructe! – VSO
SA-to me bring John some fruits

Let John bring me fruits!

b. Ion *să-mi aducă* nişte fructe! – SVO

c. **Să-mi aducă* nişte fructe Ion! – VOS

The sentences in (5) express an order at the third person, sg. A lexical subject can appear in this context, either after or before the subjunctive verb : (5a) vs. (5b). The order with a subject following the direct object NP (5c) gives an ill-formed construction. This situation corresponds to the general pattern of Romanian clauses, allowing for a free alternation SVO/VSO, but ruling out VOS (cf. Motapanyane (1989)).

In root clauses, there are other constituents, besides the subject, that could precede the subjunctive verb. e.g. :

(6) a. Mîine Maria *să* *vină* mai devreme!
tomorrow Mary SA come more early
Tomorrow Mary should come earlier.

b. Pe Ion *să* *nu-l* *mai* *invite*!
pe-John SA not-him more invite
John, don't invite him any more.

The paradigm in (6) shows that adverbs (6a) or other X-phrases (6b) can be dislocated in front of the subjunctive verb. This word order also follows the pattern of indicative clauses.

1.2. ADJUNCT CLAUSES

Subjunctive clauses function as complements of purpose, time, consequence, manner and so on. In these contexts, the complementizer (COMP) *ca* 'that' or a preposition may precede the subjunctive clause. Another option consists in the direct selection of the subjunctive, without 'introductory' element :

(7) a. pro Intru încet ca *să* *nu* *se* *trezească* Maria.–VS
(I) enter slowly that SA not REFL wake up Mary
I come in slowly so that Mary does not wake up.

b. pro Intru încet ca Maria *să* *nu* *se* *trezească* – SV

c. pro Am plecat fără *să* *mă* *vadă* mama. – pro
(I) have left without SA me see mother
I left without my mother seeing me.

d. *pro Am plecat fără mama *să* *mă* *vadă*. – SV

e. pro Am venit *să* *mă* *consulte* doctorul – VS
(I) have come SA me check doctor-the
I came to be checked by a doctor.

f. *pro Am venit doctorul *să* *mă* *consulte* – SV

The long paradigm in (7) points to the following facts : when COMP *ca* 'that' is present (7a, b) the word order can freely vary between SV/VS; the versions without COMP do not allow for SV (7f), even when a preposition selects the subjunctive clause (7d). In these cases, the obligatory order is VS.

1.3. Θ-SELECTED CLAUSES

There are many classes of verbs selecting a subjunctive clause as a direct complement. The subordinate clause can be preceded or not by COMP, as mentioned for

the adjunct contexts in (7). The restriction to VSO order in COMP-less clauses is maintained.

(8) a. pro Vreau ca Maria *să* *vină* mai devreme. - SV
(I) want that Mary SA come more early
I want Mary to come earlier.

b. ?pro Vreau ca *să* *vină* Maria mai devreme[3]. – VS

c. pro Vreau *să* *vină* Maria mai devreme. – VS
(I) want SA come Mary more early

d. *pro Vreau Maria *să* *vină* mai devreme. – SV

The restriction on word order in (8) parallels the situation in (7) : COMP allows for free SVO/VSO (8a, b), while absence of COMP triggers obligatory VSO (8c vs d).

The second peculiarity of subjunctive clauses in θ-position consists in the possibility of having a subject coreferent to the subject of the matrix clause, as mentioned under (1). This leads to a contrast between Romanian and other Romance languages, as illustrated in (3), where sentences such as (1) entail obviation.

It is clear that the restrictions on word order must be related to the lack of obviation in Romanian. Also, these facts must be reflected into the structural configuration of the subjunctive clause. In this respect, notice that the option between +COMP/-COMP clausal level exists only in Romanian, while in other Romance languages the presence of 'that' is obligatory. With three observations in mind (i.e. +/- COMP; SVO/VSO; lack of obviation), I return to subjunctive clauses in root contexts trying to give an adequate description of their internal structure.

II. THE STRUCTURE OF A SUBJUNCTIVE IP

In Pollock (1989) IP is analyzed as a complex of functional heads, standing each for an inflectional specification, e.g., T contains the morpheme of [tense], Agr brings the marks for subject-verb agreement. Each head complies with the X-bar theory and develops a maximal projection at DS. At Surface-Structure (SS) the verb 'raises' to inflection, through a movement from head-to-head, picking up the respective affixes.

I will adopt this analysis as theoretical background, but I will follow the hierarchy proposed in Belletti (1989) for Romance languages.[4] Furthermore, I will propose the extension of the IP-complex to a new functional head, corresponding to the morphology of mood (M).

Let us observe first the inflectional morphology of the subjunctive verb in Romanian. e.g. : *să vină* 'come-SUBJUNCT, 3rd, sg.'. There is *să*, an unbound morpheme marking the subjunctive mood. The verb shows a suffix for person/number agreement with the subject. There is no visible mark for [tense] features. But a specification for [aspect] must be present because there is a parallel form of subjunctive in Romanian, used in contexts of 'consecutio temporum', e.g.: *să vină* '(he/she) come' vs. *să fi venit* '(he/she) have come-SUBJUNCT'.

With respect to other elements contained in the verbal inflection, i.e. negation and clitics, the linear order shows the following : *să nu-l mai inviți* 'SA not-him more invite/SUBJUNCT'. The negation and the clitic pronoun follow the subjunctive marker *să*, and precede the adverb of negation *mai*, equivalent to Fr. *plus*, It. *piu*, Sp. *mas*, Port. *mais*. In this order, negation and clitics are separated from the verb, that remains lower, under the adverb of negation[5]. This situation had led in previous analyses (cf. Motapanyane (1990)) to propose an independent functional Agr-type head of that will host the negation and the clitics. The negation will be analyzed as an independent XP inserted under Agr at DS (cf. Pollock (1989)).

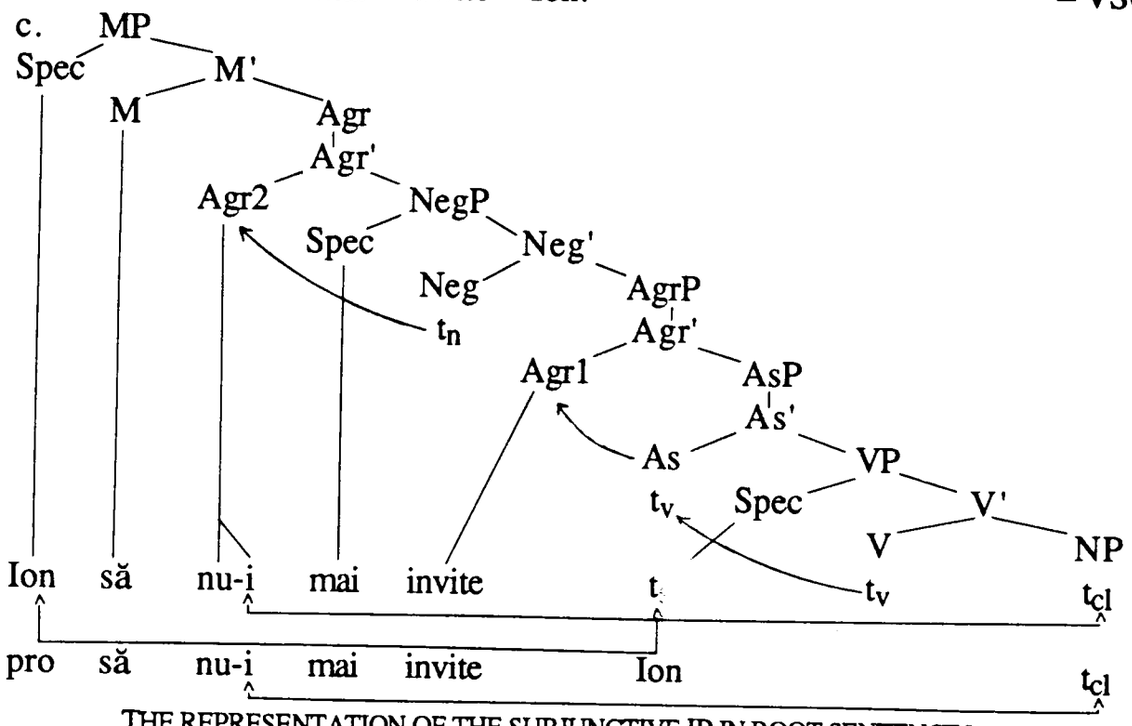
Therefore, the subjunctive IP may have three morphological features : mood, subject-agreement and aspect, to which is added a functional head specified for negation and clitic climbing. The structure of a subjunctive IP will have the representation in (9):

(9) a. Ion *să* *nu-i* *mai* *invite!* – SVO
John SA not-them more invite.

John shouldn't invite them any more.

b. *Să* *nu-i* *mai* *invite* Ion! – VSO

c.



THE REPRESENTATION OF THE SUBJUNCTIVE IP IN ROOT SENTENCES

What are the movements taking place to or within the subjunctive inflection at SS, as represented in (9)?

The verb 'raises' from V to As (presumably containing a default mark) and then to Agr1, where it receives the suffix for subject-agreement. The mood marker *să* is an unbound morpheme and as such, it cannot trigger verb-raising to its level. Therefore, the verb remains in Agr1.

The negation, under Neg at DS, raises to Agr2. The adverb of negation *mai*, base generated in SpecNegP, will follow the raised negation in linear order. The clitic pronoun *-i* 'them', generated under complement NP, climbs on Agr2 at SS as well.

With respect to subjects, I adopt two theories relevant to Romanian structures : (i) *pro* is the category corresponding to the non-lexical subject (cf. Rizzi (1986)); (ii) the subject θ-role is assigned to SpecVP (cf. Koopman (1988), Sportiche (1988)).

The theory of *pro* (cf. Rizzi (1982), (1986)) stated that an empty category with [+pronominal, -anaphoric] features is licensed in the subject position in languages where the subject can be 'understood' via the morphologic agreement on the verb. The content of this empty category is recovered through the inflectional subject-agreement. The slot for *pro* licensing corresponds to the structural subject position projected at DS through compliance with the Extended Projection Principle (EPP). EPP requires that every clause must have a subject position, in its Spec. In (9c), where the clausal level is MP, the structural subject position would be in SpecMP. In that slot, a lexical subject (9a) can alternate with a referential *pro*.

The second theory on subjects concerns the assignment of θ-role at DS. According to Koopman (1988) and Sportiche (1988), the subject θ-role must be assigned inside the maximal projection of V, namely to SpecVP. The lexical NP will be inserted in SpecVP at DS. In the derivation of SS the structure must comply with the Case theory, requiring that every lexical NP be marked for case. Therefore, the subject NP in SpecVP must receive Case at SS. If Case-assignment fails to take place in SpecVP, the lexical NP has to 'raise' at SS to a position where case is available.

In Romanian, Nom-case is assigned by verbal inflection. The mechanism of case-assignment takes place under government, in two ways : (i) under Spec-head structural agreement; (ii) under a head-complement relation. The first possibility is implemented by the highest functional head, on the clausal subject position, i.e., M assigns Nom to SpecMP in (9c), the line describing (9a). The second possibility arises from government by the lowest functional head on VP, and therefore on the adjacent SpecVP. In (9c) this last situation describes the VSO order, as in (9b).

The analysis of subject positions as proposed above accounts for the optional SVO/VSO order in a Romanian IP : two subject positions are projected at DS, a structural position (for compliance with EPP) and a θ (Thematic) position (for compliance with the θ-theory). Both positions can receive Nom-case at SS. Therefore, a lexical NP, base generated in SpecVP, can optionally raise to SpecMP (leaving an anaphoric trace/NP-t in SpecVP). The resulting order is SVO. The same NP may remain in its base position, and the resulting order is VSO. In this last configuration, the structural subject position (i.e. SpecMP) must be filled in by an expletive element, as a structure preserving condition. The expletive in Romanian (i.e. in any pro-drop language) is *pro*, as indicated in the second line under (9c).

Assuming that (9c) is the adequate representation of a subjunctive clause in Romanian, it follows that SVO/VSO must be always optional in this type of clause. Also, one would expect that adjunction of various constituents to MP be permitted, as illustrated under (6).

However, empirical data showed that this is not always true. Namely, subordinate subjunctive clauses without COMP allow only for VSO, as seen in (7) and (8). Then, what are the factors that restrict the free alternation of word order in a subordinate clause? I try to answer this question in the next section, where the representation in (9c) will be reconsidered in the context of hierarchical relations.

III. THE STRUCTURE OF A SUBJUNCTIVE COMPLEMENT IP

Let us resume the paradigm illustrating the word order in subjunctive complement clauses :

(8) a. pro Vreau ca Maria *să* *vină* mai devreme. - SV
(I) want that Mary SA come more early
I want Mary to come earlier.

b. ?pro Vreau ca *să* *vină* Maria mai devreme. – VS

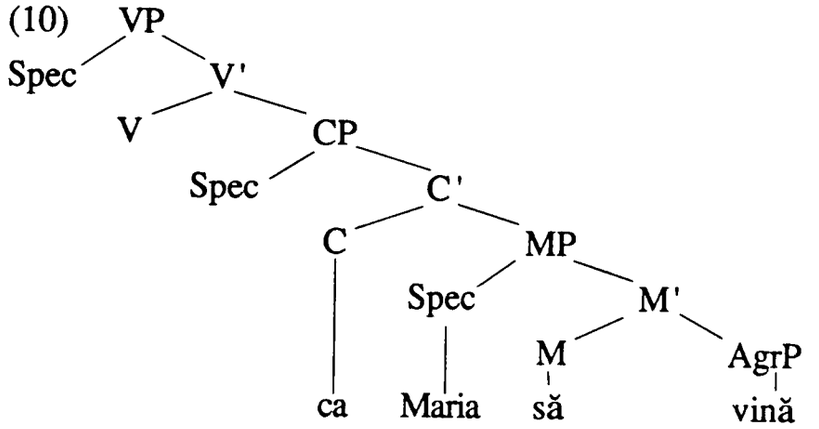
c. pro Vreau *să* *vină* Maria mai devreme. – VS
(I) want SA come Mary more early

d. *pro Vreau Maria *să* *vină* mai devreme. – SV

There are two ways of inserting a subjunctive clause in object position : either by using a COMP *ca* 'that' (8a, b) or without any other element in front of the subjunctive verb (8c, d). This last version allows only for VSO order (8c vs. 8d). The contrast between the two types of clauses stems from the presence vs. absence of COMP. How can the absence of this element determine specific changes in the internal structure of the embedded clause (i.e., in its word order) ?

The answer concerns the rules of government relations in an hierarchical structure. Consider (9c) in the context of subordination : the matrix verb (V) selects a CP, whose head C (i.e. *ca* 'that') functionally selects MP, as in (10) :

(10)



– SVO (8a)

ca pro să vină Maria – VSO (8b)

THE REPRESENTATION OF THE SUBJUNCTIVE CP IN OBJECT POSITION

The order of phrasal categories in (10) maintains the representation in (9c) : In the SVO version, corresponding to (8a), the preverbal subject occupies SpecMP. In the VSO version, illustrated in (8b), an expletive *pro* fills in the structural subject position (i.e. SpecMP) while the subject remains in its θ-position, inside VP. I claim, following Rivero (1987), that the order in (9c) is maintained due to the presence of *ca*. This lexical C ensures exclusive government of the structural subject position by the subjunctive inflection. That is, in our case, SpecMP is exclusively governed by M.

I will describe the government relations in (10) along the definition of the Relativized Minimality (cf. Rizzi (1989)) :

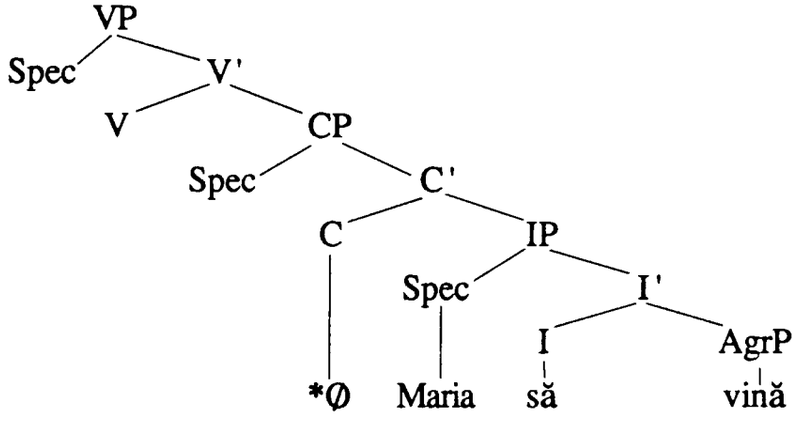
(11) X a-governs Y only if there is no Z such that

(i) Z is a Typical Potential a-governor for Y, and

(ii) Z c-commands Y and does not c-command X.

The definition in (11) discriminates between potential governors according to their hierarchy in the structure. Applying this definition to (10) we obtain the following result : matrix V θ-marks CP and can govern into IP, its direct complement; C is a lexical head of the same type as V; therefore C qualifies as a potential governor of IP. Although *ca* in C does not present inherent governing properties, it can prevent government of V into IP by the simple fact of intervening, hierarchically, between V and IP[6]. Thus, the subject position, SpecMP, will be governed only by its own head. In Rivero (1987) a similar approach led to the conclusion that a structural subject position in clauses without COMP will be excluded on grounds of double government. I will illustrate the arguments in Rivero through the representation in (12). The subjunctive clause is described as a CP-IP in Rivero (op.cit.), therefore I replace MP in (9) and (10) with IP in (12) :

(12)



PROVISIONAL REPRESENTATION OF THE COMP-LESS SUBJUNCTIVE CLAUSE : 1ST VERSION

In (12) C exists but is not lexically realized. An empty C does not enter in competition for the government of IP. On the other hand, for Rivero (1987) C must enter an 'agreement' relation with : (i) the mood marker it functionally selects; (ii) the matrix verb, under a θ-marking relation. As a result, V, C and I carry the same structural index and there cannot be any 'barrier' for government from V into IP.

The analysis in Rivero (op.cit.) concludes that a configuration like (12) exposes the structural subject position (SpecIP in Rivero (1987) and SpecMP in (10)) to double government : (i) from I under Spec-head 'agreement'; (ii) from V, under a head-complement relation. A lexical subject in this position would violate the conditions on strict mutual relation between an argument and its selecting head. This justifies the ungrammaticality of (8d) and predicts that VSO is the only possible order in such configurations.

However, Rivero's (1987) analysis, indirectly, allows for null subjects (i.e. referential *pro*) to be licensed in SpecIP. This would contradict both the theory of *pro* and the binding principles. First, referential *pro* must be licensed by a head-governor : in (12) there are two possible governors for SpecIP (V and I). Second, supposing that referential *pro* fills in SpecIP, it will be bound in the wrong domain by V, i.e., in the same governing category with the matrix inflection. Therefore, a violation of binding Principle B will follow.

Besides the theoretical points presented in the above paragraph, there are also empirical arguments against the analysis in (12). In the paradigm of root sentences in (6), it was mentioned that other elements can precede the subjunctive verb : adverbs or dislocated NPs. The same order cannot be obtained in subjunctive COMP-less clauses. e.g. :

(13) a. *pro Vreau mîine să vină Maria.
(I) want tomorrow SA come Mary

b. *pro Vreau pe Ion să nu-l mai inviţi.
(I) want PE- John SA not-him more invite (you)

c. *pro Vreau Maria să vină.
(I) want Mary SA come

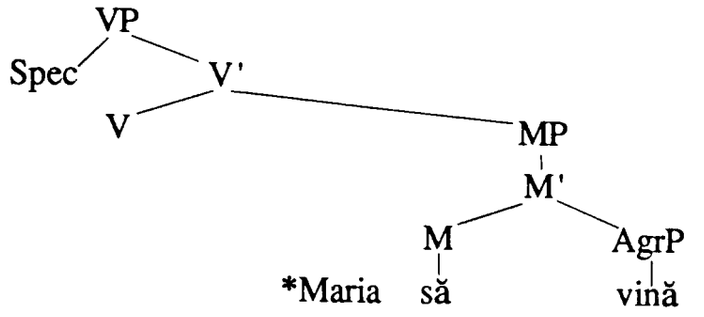
In (13) the adverb (13a) and the dislocated NP (13b) are ruled out on a par with the preverbal subject (13c). The structure in (12) cannot account for this phenomenon; on the contrary, adjunction to IP or dislocation into an A'-SpecIP must be allowed in that configuration.

I will argue that the theoretical problem arising from the existence of referential *pro* as well as the phenomena in (13) can be accounted for if the representation for (8c, d) does not present at all a CP level. That is, the matrix verb V θ-selects a subjunctive IP (= MP).

Direct embedding of IP in object position excludes the possibility of adjunction, as illustrated in (13). According to Chomsky (1986) adjunction to arguments interferes in the θ- relation between the selecting head and its object position.

With respect to the structural subject position in SpecIP (= SpecMP), I adopt the analysis of double government as proposed in Rivero (1987). But I consider that such a position must be completely excluded at DS, where the hierarchy of the clause is determined. That is, I will propose a representation where a SpecMP position is not projected at all. Let us consider a second possibility to describe the clauses in (8c, d) :

(14)



– SV (8d)

să vină Maria – VS (8c)

PROVISIONAL REPRESENTATION OF THE COMP-LESS SUBJUNCTIVE CLAUSE : 2ND VERSION

The representation in (14) accommodates two empirical facts : lack of adjunction to MP (=IP); complete exclusion of a subject position from a context of double government. This last point correctly treats lexical NPs on a par with referential *pro*.

On the other hand, the representation in (14) fails to : (i) conform with EPP; (ii) account for null subject clauses. The first point refers to the obligatory condition on structures requiring that every clause project a subject position. (14) must have a clausal subject position in order to qualify as an adequate description of an IP-structure.

The second point is also crucial to the present analysis, because subjunctive clauses such as (8c, d) not only have null subjects, but these subjects freely corefer with the subject of the matrix clause. I refer the reader to the comments on the contrast between (1) and (3) at the beginning of this paper. The observation consisted in the fact that obviation in Romanian sentences in (1) does not apply, while it is strictly observed in the 'equivalent' structures in other Romance languages (3). The representation in (14) cannot account for this contrast.

At the present stage of analysis, I have to adjust the representation in (14) so that it could cover the following factors : (i) compliance with EPP; (ii) existence of referential *pro* in IP-level subjunctives; (iii) possibility for this *pro* to corefer with the subject of the matrix clause; (iv) the same mechanism does not apply to other Romance languages.

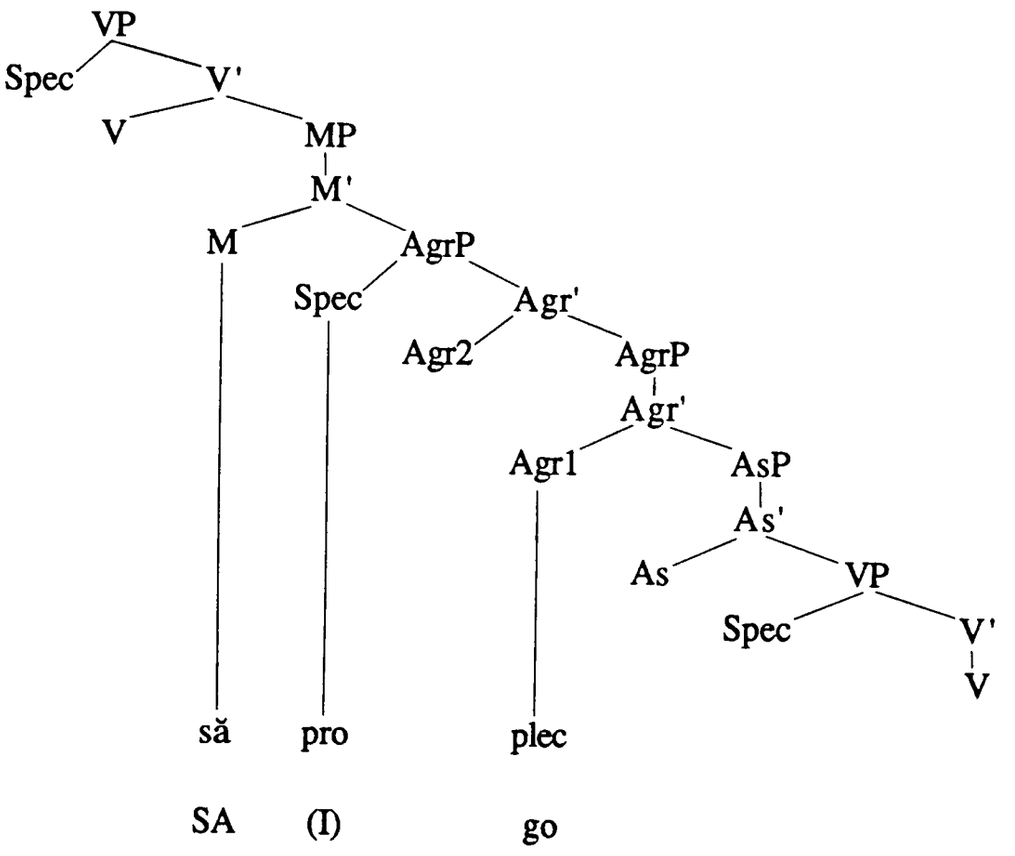
It is clear that (14) will impose structural constraints to the application of EPP. In these conditions, it is plausible to suppose that if a subject position is disallowed in the Spec of the highest inflectional head, it will be projected to the next available level, that is SpecAgrP in (9c) and (10). Let us consider the consequences of such a structure :

(15) a.

pro	Vreau	să	plec.
(I)	want	SA	go (I)

I want to go.

b.



THE REPRESENTATION OF THE IP-SUBJUNCTIVE COMPLEMENT

In (15b) EPP is complied with : a structural subject position is projected in SpecAgrP. Agr2 governs this Spec under a structural 'agreement' relation. Therefore, licensing of referential *pro* can occur, under government by Agr2. The identification of *pro* obtains, as in indicative clauses, from coindexing of Agr2 with the head carrying the subject-verb agreement marks, i.e. Agr1.

Another consequence of the analysis in (15b) consists in the fact that government from selecting V on referential *pro* is prevented. Following the conditions of Relativized Minimality (cf. (11) above) M is a lexical head, qualifying as a potential governor of AgrP. The mood mark *să* does not have governing properties, but its hierarchical position is sufficient to protect AgrP from government from a higher head (the relevant head here being matrix V). Thus, a double government relation as argued for in (12) is avoided in (15) : the subject position, situated between two governors, V and Agr2 (=I in (12)) cannot enter into a relation with V. Through Relativized Minimality Agr2 qualifies as the sole governor of the structural subject position.

The disambiguity with respect to government can also account for the possibility of having coreferent subjects in sentences like (15a). The intervening M ensures the conditions to create an opaque domain for binding at the level of the embedded clause. A pronominal element in SpecAgrP will be bound by the coreferent subject outside the domain of the matrix clause. This complies with the binding Principle B. In other words, obviation becomes irrelevant in a configuration like (15b).

One may wonder why a lexical subject cannot occupy this SpecAgrP, as to allow for SVO order in sentences such as (8d) or (15a). I will argue that this is prevented by

Case-theory. A lexical subject needs case, that is Nom. in Romanian, assigned under government by verbal inflection. As mentioned for (9c), case-assignment takes place in two ways : (i) under Spec-head government by the highest functional head; (ii) under head-complement government by the lowest functional head. In (15b) the highest functional head cannot display inflectional government and case-assignment at SS because it does not have a Spec position. This rules out SVO in (8d). Therefore, case-assignment will be possible only under the second condition, that is, through the lowest functional head on SpecVP, resulting in obligatory VSO (8c). A referential *pro* does not need case, so it can appear in a non-initial clausal subject position (15a). Finally, I should mention that VSO order implies an expletive *pro* in SpecAgrP, to fill in this argumental position and thus ensure the conditions for the preservation of the structure.

So far, I have argued that the two types of subjunctive clauses in (8) differ with respect to their underlying structures. The CP-level subjunctive received the description in (10), while the IP-level subjunctive was ultimately represented in (15b). The CP-subjunctive allows for optional SVO/VSO; the IP-subjunctive structurally restricts the choice to VSO when the subject is lexical. Comparing these results with the situation in Romance, as illustrated and commented under (3), it follows that lack of obviation in Romanian is due to the possibility of having IP-level subjunctives. In Romance, a lexical COMP is obligatory, therefore their underlying structure would correspond uniquely to (10). However, there are empirical data that would contradict such a clear-cut distinction between Romanian and Romance. In the next section, I propose a closer look at the situation of subjunctive complements in other languages.

IV. ROMANCE AND GREEK SUBJUNCTIVE COMPLEMENTS

Romanian structures in (8c) and (15b) have been compared with Greek equivalent sentences (cf. Rivero (1987), Dobrovie-Sorin (1987), Kempchinsky (1987)). The general tendency consisted in justifying (i) the phenomenon of replacing the infinitive with the subjunctive, (ii) the existence of a morphological mood marker, i.e. *să* and (iii) the existence of COMP-less subjunctive clauses, by the linguistic environment within the Balkan peninsula.

While this comparative approach is justified, I would also like to draw attention to a few facts that show a fundamental distinction between Greek and Romanian with respect to the subjunctive syntax.

The IP-subjunctive in Greek displays a paradigm that cannot be reduced to the representation in (15b). e.g. :

(16) a. thelo kathe anthropos na ksekurasti
(I) want every person (NOM) NA rest-(he)

I want every person to rest.[7]

b. thelo ton Yani na fivi
(I) want the John(ACC) NA leave

I want John to leave.

c. thelo na pao stin Athina
(I) want NA go-(I) to Athens
I want to go to Athens

d. *thelo oti kathe anthropos na ksekurasti.
(I) want that every person NA rest -(he)

The first remark on (16) : in Greek a CP-level subjunctive is not possible at all (16d). This marks a first important contrast with Romanian, where CP-subjunctive clauses are common (cf. (8a, b)). In (16c) the subjects of the two clauses corefer, and the structure is grammatical. This phenomenon looks similar to lack of obviation in sentences such as (1) and (15a). However, the underlying structures of (16c) and (15a) cannot be the same. This indication comes from the word order in the first two examples : in (16a) the subject in Nom. precedes the subjunctive marker; in (16b) the same NP carries Acc-case. (16a) contrasts with (15b) in the choice of word order : in Romanian VSO is obligatory, while in Greek SVO seems normal. The possibility of an Acc-marked NP in (16b) increases the structural gap between Greek and Romanian : in Greek the matrix V either selected two objects or it selected only one object, i.e., the subjunctive clause, whose subject position is case-marked under government from matrix V. Indeed, according to Felix (1989), the structure in (16b) shows Exceptional Case Marking (ECM). There is no equivalent construction in Romanian.

On the contrary, examples of dialectal or archaic use of subjunctive complements in Romance would rather lead to a structural relation between these languages and Romanian. Consider the following paradigm :

(17) a. pro Non sapevo fosse malato.[8]
(I) not knew be (he) sick
I didn't know him to be sick.

b. pro Credo sia arrivato Piero.
(I) think be arrived Peter
I think that Peter has arrived.

c. *pro Non sapevo Giorgio fosse malato.

d. *pro Credo Piero sia arrivato.

The Italian subjunctive clauses in (17) lack a complementizer. The subjunctive does not present an inflectional marker in this language, but the structural restrictions in an embedded context look similar to Romanian in (8c, d). Thus, a lexical subject can appear only after the verb, that is VSO is obligatory (17b) vs. (17c, d). Under the same conditions, a referential *pro* is possible in the subjunctive clause (17a). Therefore, it is plausible to say that the underlying structure of (17) is the same as (15b).

The syntactic comparison proposed in this section leads to the following distinction: (i) IP-subjunctive clauses do not constitute a peculiarity of Romanian grammar: they appear in Romance languages (in dialectal or marked constructions) as well and they display the

same restrictions on word order. (ii) Greek IP-subjunctive clauses have a fundamentally different syntax, despite resemblences in inflectional morphology and spread use.

The conclusions to be drawn on the basis of these considerations would consist in two general remarks : (i) at a certain point of its history, Romanian grammar opted for the spread use of subjunctive clauses under the influence of similar phenomena in Balkan languages; (ii) this option was possible because there existed already a pattern for two subjunctive clause types in Romance: other Romance languages have dropped the IP-level pattern, while Romanian maintained it and developed it under the factors in (i).

FOOTNOTES

1 Example quoted from Kempchinsky (1987).

2 The infinitive still occurs in Romanian, even in complement positions, but with a very reduced number of verbs; e.g.:

(i) modal *putea* 'can'

(ii) aspectual *a începe* 'to begin'

(iii) raising verb *a părea* 'to seem'

3 The use of adjacent *ca să* is subject to idiolectic variation. There are speakers who prefer to have a lexical element between *ca* and *să* (as in (8a)).

4 In Pollock (1989) T is higher than Agr in the structural representation. In Belletti (1989) this order is reversed for Romance, where empirical evidence shows that the mark of [tense] is closer to the verbal root than the suffix of subject- agreement.

5 This order is different from other Romance languages, where negation and clitics immediately precede the inflected verb, while the adverb of negation remains lower than the verb. e.g.:

(i) Je ne l' invite plus.
I not him invite more

6 The exclusion of V in terms of hierarchical government does not suffice when it comes to the delimitation of binding domains. It seems that COMP can prevent government of V on SpecIP (here SpecMP) but it cannot avoid wrong binding of the pronominal subjects in this position. Hence the obviation condition, as illustrated in (3a, c).

7 The Greek examples are quoted from Felix (1989).

8 All the Italian examples are quoted from Cinque (1981).

REFERENCES

Belletti, A. 1989. 'Agreement and Case in Past-Participial Clauses in Italian.' ms. Geneva : Université de Genève.

Chomsky, N. 1981 . *Lectures on Government and Binding*. Dordrecht : Foris.

Chomsky, N. 1986. *Barriers*. Cambridge, Mass : MIT Press.

Cinque, G. 1981. 'On Keenan and Comrie's Primary Relativization Constraint.' *Linguistic Inquiry* 12:2 : 293-308.

Coteanu, I. 1981. *Structura si evolutia limbii romane (De la origini pina la 1860)*. Bucuresti : Editura Academiei.

Dobrovie-Sorin, C. 1987. *Syntaxe du roumain. Chaînes thématiques*. Thèse de Doctorat d'Etat. Paris : Université de Paris 7.

Dobrovie-Sorin, C. 1990. *The Syntax of Romanian*. Dordrecht : Foris.

Farkas, D. 1982. *Intensionality and Romance subjunctive relatives*. Indiana, Bloomington: IULC.

Felix, S. 1989. 'Finite Infinitives in Modern Greek.' In Bhatt, C., E. Lobel & C. Schmidt (eds.), *Syntactic Phrase Structure Phenomena*, 113-132. Philadelphia : Benjamins.

Kempchinsky, P. 1987. 'Operators and Binding in Subjunctive Clauses.' ms. Los Angeles: UCLA.

Koopman, H. 1988. 'La structure interne du VP en Hollandais.' Talk at Séminaire interdépartemental de recherche en linguistique, Geneva : Université de Genève.

Kuroda, S-Y. 1986. 'Whether we agree or not : Rough ideas about the comparative syntax of English and Japanese.' ms. USCD.

Motapanyane, V. 1989. 'La position du sujet dans une langue à l'ordre SVO/VSO.' *Rivista di Grammatica Generativa* 14 : 75-103.

Motapanyane, V. 1990. *Theoretical Implications of the Sentential Complementation in Romanian* (forthcoming). Thèse de Doctorat. Geneva : Université de Genève.

Pollock, J-Y. 1989. 'Verb Movement, Universal Grammar and the structure of IP.' *Linguistic Inquiry* 20:3 : 365-424.

Rivero, M-L. 1987. 'Barriers and Rumanian.' In de Cesaris & C. Kirschoner (eds) *Proceedings of the Linguistic Symposium on Romance Languages XVII*. Amsterdam : Benjamins.

Rizzi, L. 1982. *Issues in Italian Syntax*. Dordrecht : Foris.

Rizzi, L. 1986. 'Null Objects in Italian and the Theory of pro.' *Linguistic Inquiry*, 17:3 : 501-557.

Rizzi, L. 1989. *Relativized Minimality*. Cambridge, Mass : MIT Press.

Sportiche, D. 1988. 'A Theory of Floating Quantifiers and Its Corollaries for Constituent Structure.' *Linguistic Inquiry*, 19:3 : 425-449.

LE DISCOURS FRANÇAIS: VERS UNE NÉGATIVISATION ACCRUE

Suzanne Pons-Ridler
University of New Brunswick

Geneviève Quillard
Royal Military College of Canada

RÉSUMÉ

Les données fournies par le Dictionnaire des fréquences *montrent que, de la Révolution à l'époque récente, s'accroît progressivement l'utilisation des formes négatives en français. D'autre part, une étude comparative chiffrée de textes traduits met en évidence la tendance marquée des francophones à préférer dans de très nombreux cas l'approche négative, alors que l'anglophone choisit l'affirmative.*

La présentation de données chiffrées et d'exemples variés montrera que ce phénomème est extrêmement répandu et atteint tous les niveaux du discours. Il semble de plus continuer à s'étendre dans bien des régions francophones.

Pour expliquer l'origine de cette négativisation du discours français, nous émettrons un certain nombre d'hypothèses, d'ordre socio-politique, culturel et linguistique.

INTRODUCTION

Cette étude de la négation portera en premier lieu sur des données chiffrées publiées dans deux dictionnaires de fréquences, celui de Brunet: *Le Vocabulaire français de 1789 à nos jours* (1981), et celui de Johanson et Hofland: *Frequency Analysis of English Vocabularly and Grammar* (1989).

Seront présentés ensuite les premiers résultats d'une comparaison de magazines et documents bilingues qui permettent d'analyser de façon plus précise les différences d'approche des deux langues.

Dans une deuxième partie, l'étude d'exemples précis montrera à quel point l'ensemble du discours français paraît marqué par la négation quand on le compare à l'anglais.

Dans une troisième section cette tendance au mouvement ascendant de la négation en français sera exemplifiée.

Enfin une série d'hypothèses socio-historiques, culturelles et linguistiques seront proposées, hypothèses susceptibles d'éclairer certains aspects de ce phémonène.

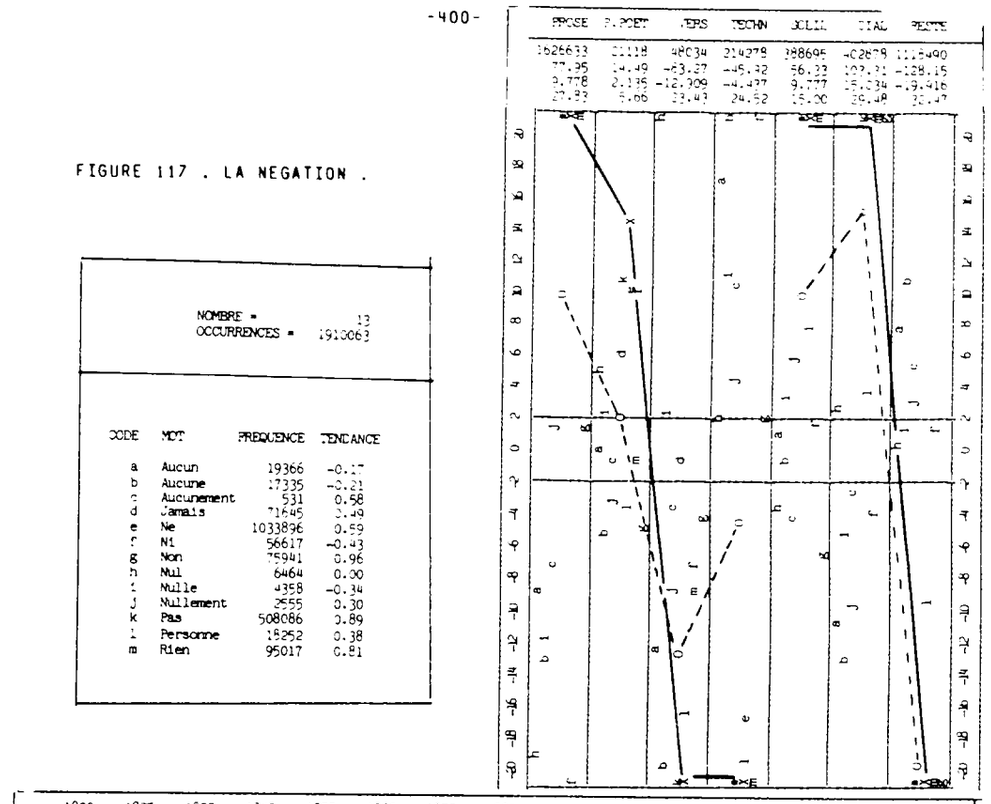
NÉGATION EN FRANÇAIS: STATISTIQUES D'APRES BRUNET

Dès 1981, dans son analyse du *Vocabulaire français de 1789 à nos jours*, faite à partir des données du TRÉSOR DE LA LANGUE FRANÇAISE, Etienne Brunet constatait: "La courbe de la **négation** (graphique 117) montre une rupture entière entre les deux siècles. Toutes les tranches du XIX^{e} sont largement déficitaires et toutes celles du XX^{e} largement excédentaires (r = 0,76). Bien entendu la catégorie est dominée par le poids de la négation **ne** qui accapare plus de la moitié des occurrences. Mais la catégorie est homogène: seuls s'écartent de la ligne les adjectifs négatifs (**aucun** et **nul**) et le coordonnant **ni** (qui régresse comme les autres coordonnants)." (p. 339).

Les chiffres qui ont permis d'établir ces courbes et celles qui sont présentées ultérieurement sont présentés en note dans un tableau récapitulatif reconstitué.[1]

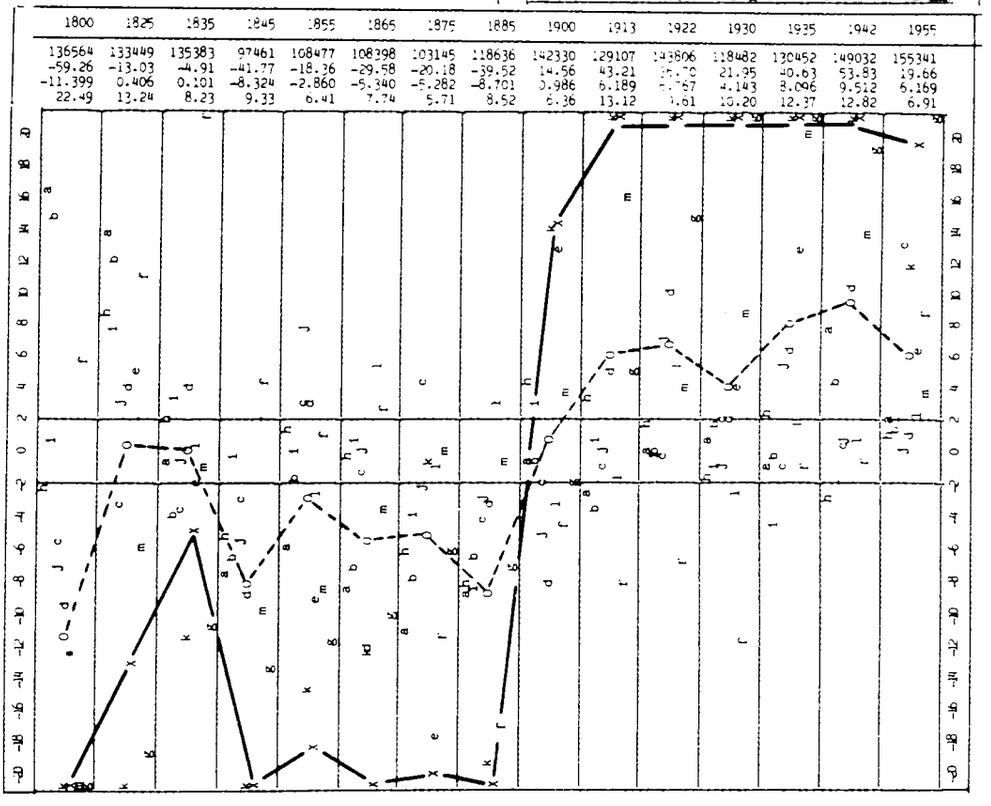
Si l'on considère les courbes des quatre négatifs les plus fréquents, 'ne', 'pas', 'jamais' et 'rien', on voit apparaître un schéma tout à fait similaire, avec des montées et descentes au cours du 19^{e} siècle qui correspondent dans le cas de 'jamais' et 'rien', ou se complètent pour 'ne' et 'pas'. En effet dans ce dernier cas, à la chute de l'emploi de 'ne' en début de siècle, correspond une montée de 'pas' très nette malgré quelques fluctuations. C'est l'époque où la négation à deux termes s'établit fermement. Dès le 20^{e} siècle, à la montée en flèche de 'ne' et 'pas', correspond une courbe moins ascendante pour 'jamais' et 'rien', mais tout de même significative. A l'exception de 'pas', toutes ces formes marquent une déclivité notable aux alentours des années 1930, ce qui pourrait s'expliquer par le contexte socio-politique de cette période.

FIGURE 117 . LA NEGATION .

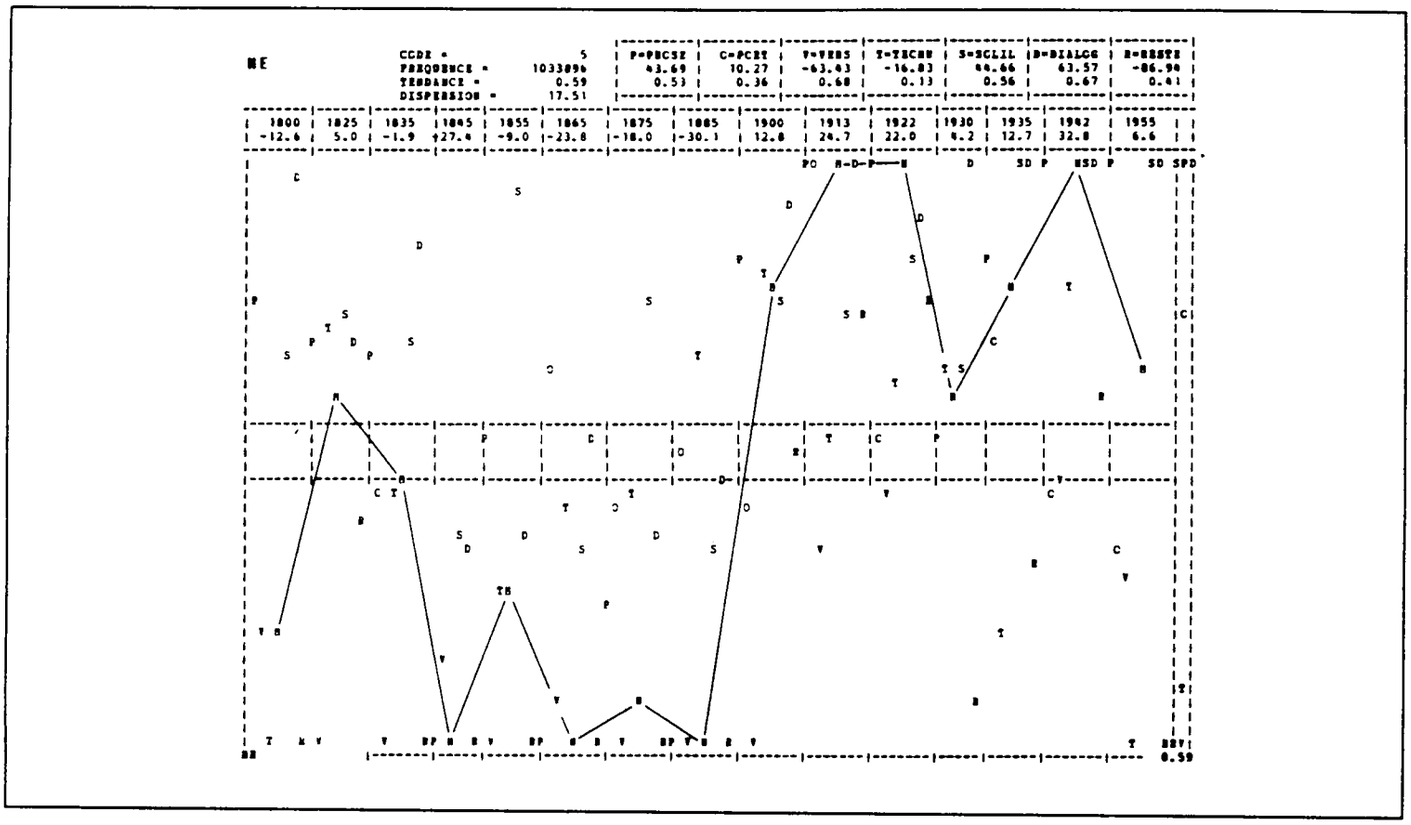


NOMBRE =	13
OCCURRENCES =	1910063

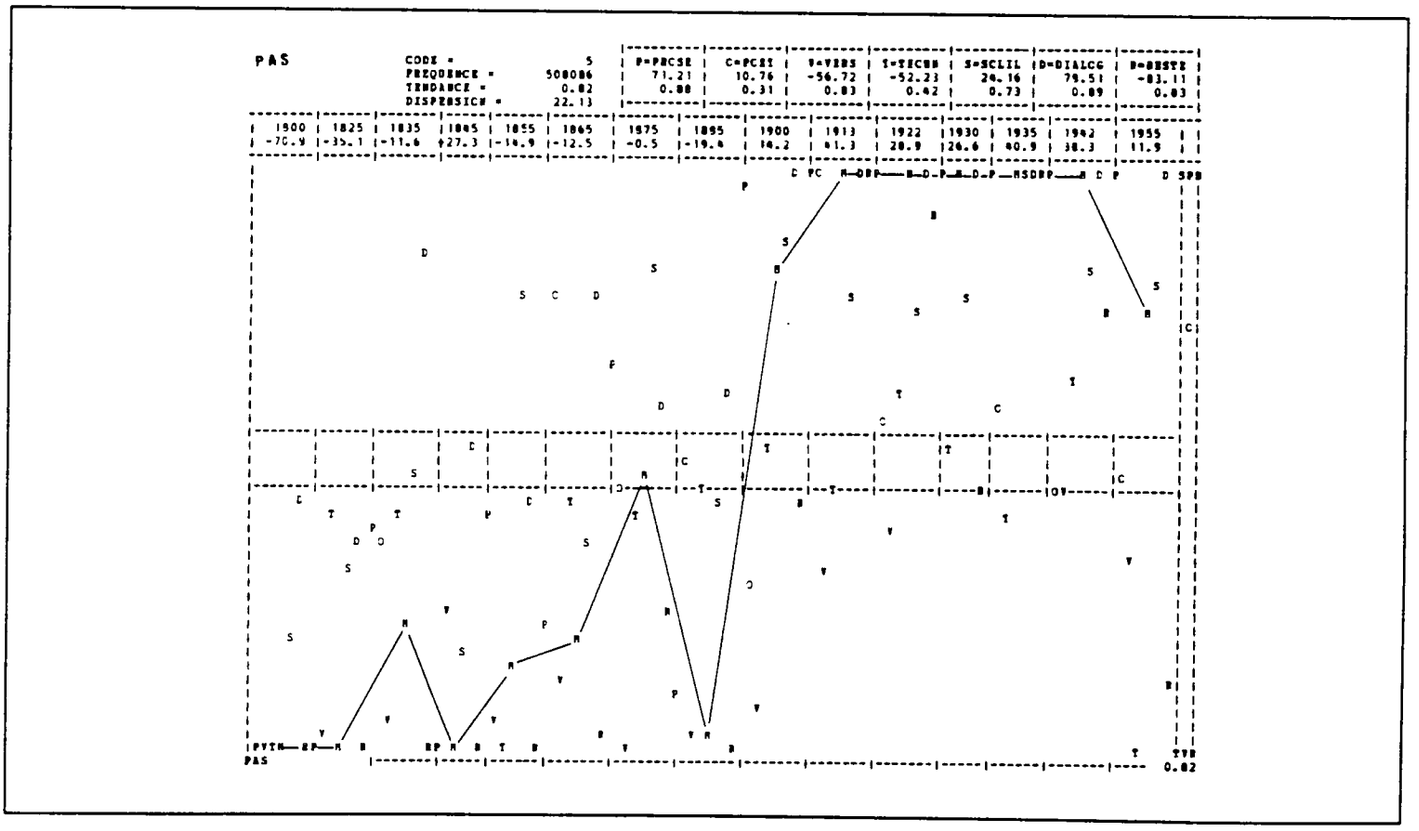
CODE	MOT	FREQUENCE	TENDANCE
a	Aucun	19366	-0.17
b	Aucune	17335	-0.21
c	Aucunement	531	0.58
d	Jamais	71645	0.49
e	Ne	1033896	0.59
f	Ni	56617	-0.43
g	Non	75941	0.96
h	Nul	6464	0.00
i	Nulle	4358	-0.34
j	Nullement	2555	0.30
k	Pas	508086	0.89
l	Personne	18252	0.38
m	Rien	95017	0.91



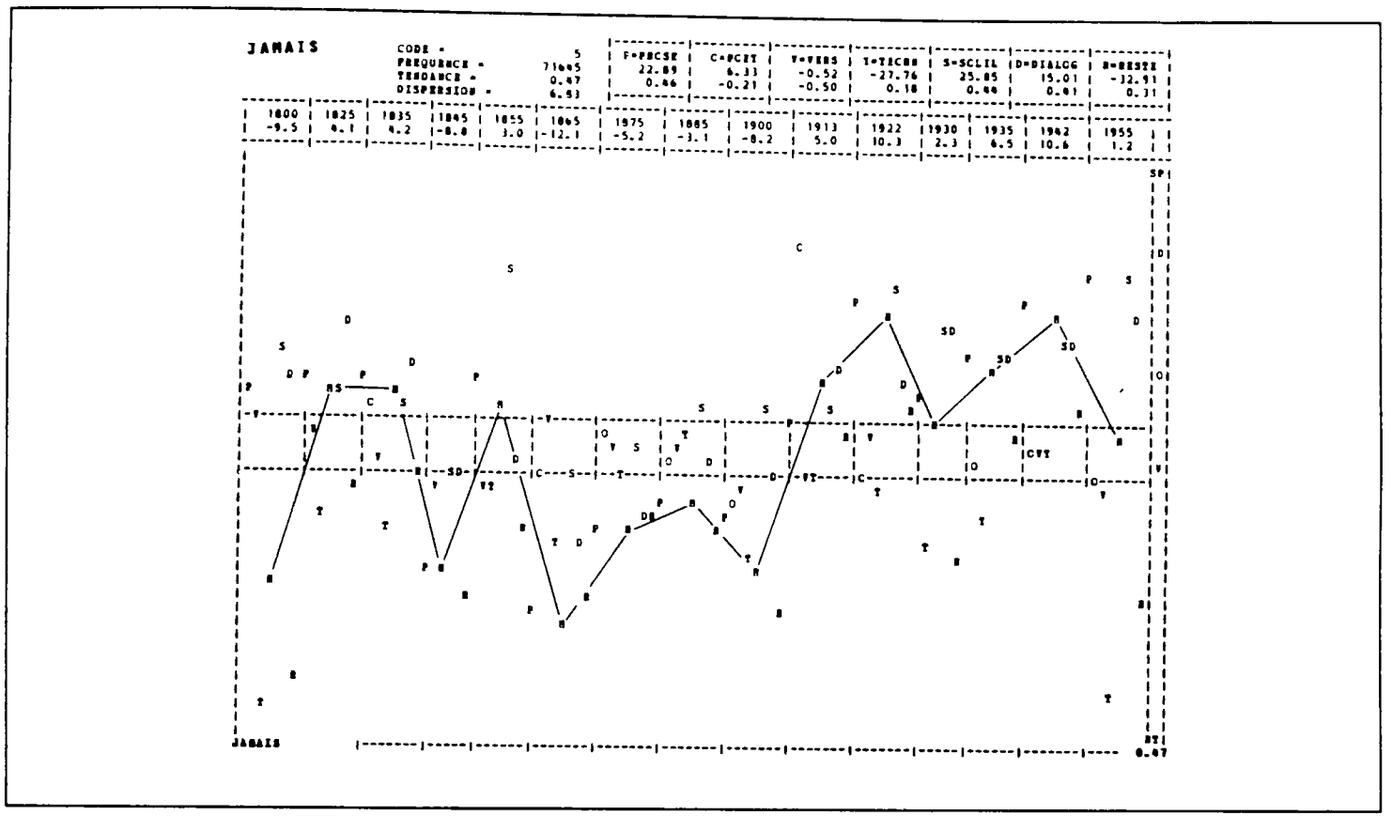
LA COURBE DE LA NÉGATION D'APRÈS BRUNET



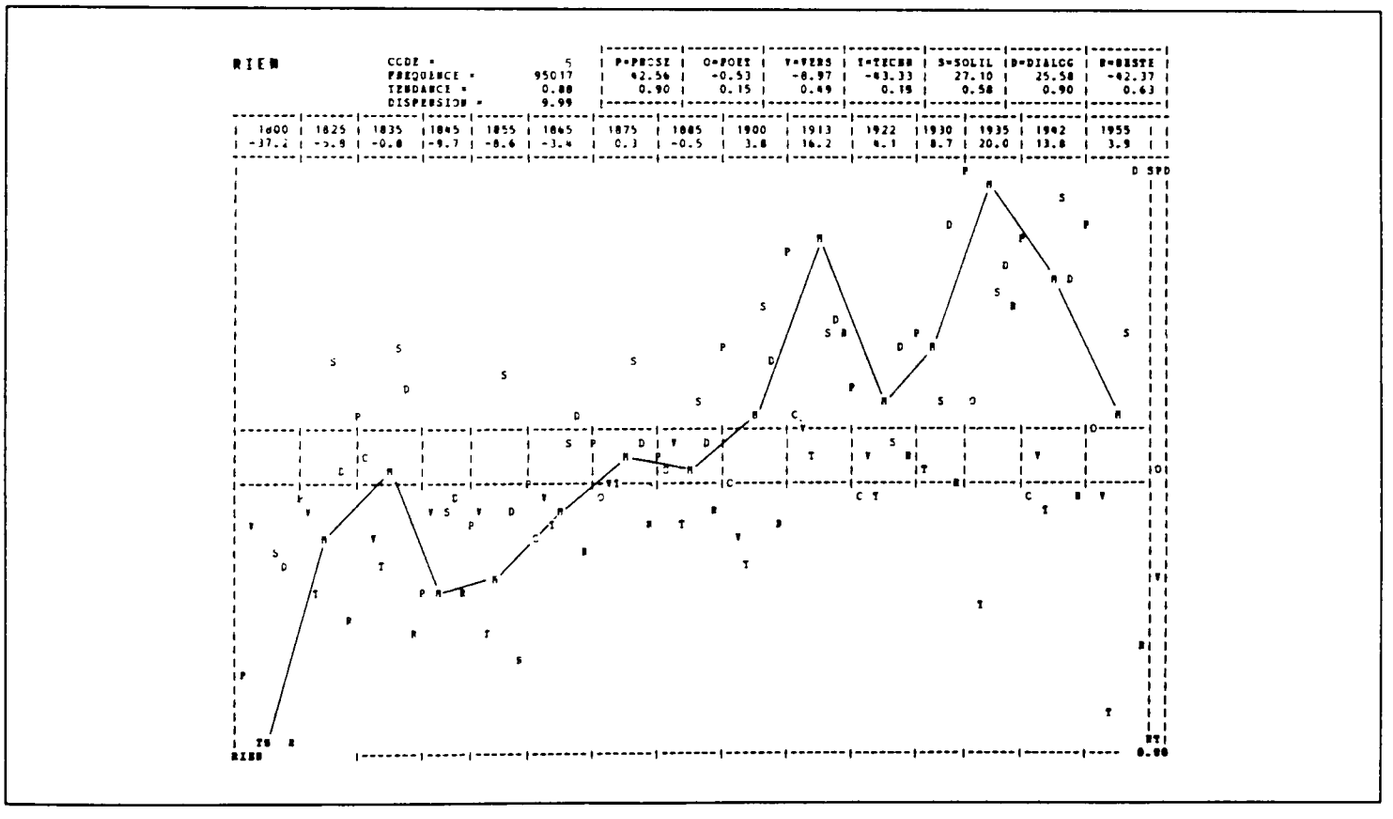
LA COURBE DE *NE* (BRUNET, VOL. 3: 269)



LA COURBE DE *PAS* (BRUNET VOL. 3: 299)



LA COURBE DE *JAMAIS* (BRUNET VOL. 3: 211)



LA COURBE DE *RIEN* (BRUNET VOL.3: 365)

NÉGATION EN ANGLAIS : STATISTIQUES D'APRÈS JOHANSSON ET HOFLAND

Ces chercheurs effectuent une analyse approfondie de la fréquence des mots selon leur qualité ou fonction grammaticale et selon les genres littéraires étudiés. Cependant ils ne traitent pas de la négation en tant que telle et ne répertorient pas historiquement le vocabulaire : ils ne considèrent que l'état présent. Il est donc difficile de comparer les statistiques de cette étude avec celles de la précédente. Dans l'introduction cependant figure le rang des 50 mots les plus utilisés de la langue anglaise : 'not' occupe la 23e place et 'no' la 47e (original LOB corpus, p. 19). Le *Dictionnaire des fréquences* du TRÉSOR DE LA LANGUE FRANÇAISE (sur lequel Brunet base son analyse) donne pour la période la plus récente 'ne' au 12e rang, 'pas' au 22e.[2] 'No' correspondant à la fois à 'aucun', 'aucune', 'nul', 'nulle' et 'non', il n'est pas facile de déterminer avec exactitude le rang correspondant. On peut noter toutefois que 'ne' et 'pas' sont relativement bien plus fréquents que leur correspondant anglais 'not'.

Dans cet ouvrage, l'analyse différentielle des fréquences selon les genres est plus fine que celle de Brunet. Le texte français en distingue 7, le texte anglais 15.

Groupés sous la rubrique XNOT qui comprend 'not' et 'n't', on relève les chiffres suivants:

	Text categories	Absolute frequency	Relative frequencies in words per million
A	Press: reportage	487	5463
B	Press: editorial	424	7787
C	Press: review	232	6760
D	Religion	288	8375
E	Skills, trades and hobbies	428	5565
F	Popular lore	517	5803
G	Belles-lettres, biography,essays	1026	6605
H	Miscellaneous (government documents, foundation reports, industry reports, college catalogues, industry house organ)	339	5579
J	Learned and scientific writings	845	5219
K	General fiction	673	11367
L	Mystery and detective fiction	557	11334
M	Science fiction	104	8582
N	Adventure and western fiction	589	9917
P	Romance and love stories	783	13186
R	Humour	162	8900
A-J	Group of informative text categories	4586	6064
K-R	Group of imaginative textcategories	2768	11140

TABLEAU OBTENU À PARTIR DU TABLEAU 1, P. 2 ET TAG XNOT, P. 38

Bien que ces chiffres ne portent que sur un élément négatif 'not / n't', il s'agit du plus fréquent (23e rang du vocabulaire) et ils sont significatifs de l'ensemble. Il est donc intéressant de noter à quel point la fréquence varie selon les genres. Si l'on considère la 2e colonne (fréquence relative en mots par millions), on note une différence allant de 5219 pour la fréquence la plus basse (Learned and Scientific Writings) à 13186 pour la plus

haute (Romance and Love Stories), soit un écart de plus du double. L'ensemble des textes classés 'imaginative' comporte une fréquence de 'not / n't' quasiment deux fois plus élevée que celle des textes d'information dits 'informative' (11140 versus 6064).

Ces chiffres expliquent en partie les grands écarts d'emploi qui ont été constatés dans les dépouillements de textes variés bilingues, variation au niveau de la fréquence générale d'emploi pour un texte, et variation de langue à langue.

COMPARAISON DE TEXTES BILINGUES

Les données recueillies proviennent de magazines bilingues et de documents divers. Ils correspondent donc aux catégories A et H selon la classification de Johansson et Hofland, groupes à fréquence relativement basse (5463 et 5579).

Les semi-négatifs n'ont pas été inclus, ni les termes restrictifs comme 'sans', 'sinon', 'non seulement'. N'ont été retenus que les termes négatifs de la liste de Brunet.

		A	B	C	D	E	D+E
		Total des négations	**Négation dans les 2 langues**	**Angl = 0 / Fran = 0**	**Négations en français pas en anglais**	**Négations en anglais pas en français**	**Écart absolu**
Canadian	mai 1989 (116p)	84	38	- / 1	36	9	F = + 25
	fév.1990 (115 p)	84	29	- / 3	35	17	F = + 18
	mars 1990 (86 p)	56	15	4 / 2	27	8	F = + 19
total		224	82	4 / 6	98	34	62
En Route	fév.1989 (109 p)	102	59	- / 3	26	14	F = + 12
	mai 1989 (113 p)	94	57	1 / 2	28	6	F = + 22
	mars 1990 (121 p)	103	49	5 / 14	20	15	F = + 5
total		299	165	6 / 19	74	35	39
Empress	mars 1987 (60 p)	106	66	1 / 4	22	13	F = + 9
total		629	313	11 / 29	194	82	110

Cette comparaison permet de constater une nouvelle fois l'usage plus fréquent de la négation en français qu'en anglais. En valeur absolue, dans les textes dépouillés cette différence s'élève à 17.51%, un pourcentage tout à fait substantiel. Ce chiffre ayant été obtenu sans tenir compte ni des semi-négatifs, ni des formes restrictives du type 'sans' ou

'in-'..., on conçoit que le discours français soit perçu comme nettement plus négatif que l'anglais.

ETUDE DE CONTEXTES OU LE FRANÇAIS PRÉFÈRE UNE APPROCHE NÉGATIVE ET L'ANGLAIS UNE APPROCHE AFFIRMATIVE

Si l'on classe les exemples selon les termes grammaticaux qui donnent lieu à une forme négative en français, on constate que tous les grands groupes sont représentés: verbes, adjectifs, adverbes, noms, et même les 'modifiers', ou que la négation n'est pas présente du tout dans le texte anglais. Ceci montre à quel point le phénomène est étendu et porte véritablement sur l'ensemble du discours (même si en français le verbe est l'élément négativisé privilégié). On conçoit que l'anglophone qui utilise plus spontanément une tournure affirmative ait du mal à identifier les cas où le francophone préfère une approche négative.

Voici quelques exemples:

VERBE

- If you have what you feel are extremely compelling reasons *for revoking* your intended sabbatical... / Si des raisons impérieuses vous obligeaient à *ne pas prendre* le congé sollicité...(note de service, Glendon College, York University, Jan. 1986)

- They *lose* their trust in people... / Ils *n'ont plus confiance* en personne...(prospectus: *Wife assault : it is a crime*)

- you'll *avoid* having to make a special trip... / vous *ne serez pas obligé* de vous présenter à votre succursale...(prospectus des Obligations d'épargne du Canada, 1989)

ADJECTIF

- The Native people had a *thorough* knowledge... / Ces cours d'eau *n'ont aucun secret* pour les autochtones...(*An Informal History of the Land and Its People*, p. 7)

- The majority of the class will be *new* to RMC... / Les recrues qui, pour la plupart, *ne sont jamais venues* au RMC...(note de service, Royal Military College, 27 août 1990)

- Some *legitimate* questions remain... / Ce régime fait l'objet de contestations *non dénuées de fondement* ...(*Language and Society*, Winter 1984, p. 3)

ADVERBE

- The establishment has become *simply* by chance... / Ce bouillon de culture *n'a pas d'autres causes* que le hasard...(*Canadian*, July 1989, p. 44)

- Council *willingly* devoted unscheduled time... / le Conseil *n'a pas hésité* à y consacrer des heures non prévues àl'horaire...(*CAUT Bulletin*, June 1989, p. 3)

- deadlines *frequently* make it difficult... / *il n'est pas rare* que les échéances rendent difficile...(prospectus, *Nouveautés du Centre d'édition du gouvernement du Canada*, (D) 12/89)

NOM

- Progress became the watchword of the day... / On ne parle plus que de progrès...(An Informal History of the Land and Its People)

- it is the *centre* of the record industry... / on *ne saurait passer sous silence*...(*Canadian*, July 1989, p. 34)

- We face the day-to-day *complexities*... / Il est vrai que la vie *n'est pas facile*...(*En Route*, October 1989, p. 196)

'MODIFIER'

- *What* better time... / Il *n'est pas de* meilleure occasion... (prospectus, *Federation of Ontario Naturalists*, 28 septembre 1987)

- to Upper Canadians outside the circle... / Ceux qui ne font pas partie... (*An Informal History of the Land its Its People*)

- those *far from* railways... / Celles qui *ne sont pas desservies*...(id.)

NÉGATION ABSENTE EN ANGLAIS

- That's what makes a champion... / Ce qui fait les champions, c'est *sans aucun doute*... (*En Route*, July 1989, p. 59)

- You decide to stick a bit... / On décide de rester, *ne serait-ce que* l'espace d'un instant...(*Canadian*, October 1989, p. 40)

- .../ ce n'est que retarder l'inévitable... (*An Informal History of the Land and Its People*)

Loin de se limiter aux termes proprement négatifs, ce phénomène s'étend aux semi-négatifs et à d'autres éléments, aux charnières logiques notamment, qui contribuent à négativiser le discours français. Par ailleurs, il semblerait que cette tendance négativisante en français continue à s'accentuer.

NÉGATIVITÉ EN PROGRES

1. L'AFFIXE ANGLAIS '-FREE'

Il est intéressant de remarquer que certains termes anglais, notamment le suffixe -**free,** sont **toujours** traduits par une forme négative en français.[4] Or, cet affixe, apparu au cours de ces dernières années dans l'usage nord-américain, a été amené à remplacer un terme négatif (without) dans les contextes où l'*absence* de tel ou tel élément constituait un facteur perçu comme positif. Ceci semble indiquer que, en anglais, les changements linguistiques tendent à présenter la réalité d'une façon plus positive. Le français, lui, n'a

d'autre solution, pour traduire ce sens très positif, que d'utiliser une forme négative. En voici quelques exemples :

- The needles on the ground will eventually create a natural *carefree* carpet... / Ces arbres créent un tapis naturel qui *ne nécessite pas de soin*... (note de service, Glendon College, May 18, 1990).

- But the history of electronic test marketing has not always been *problem-free*... / L'histoire des marchés-tests électroniques n'a pas toujours été *sans problème*... (*En Route*, March 1989, p. 78).

- 40 to 45 percent remain essentially *headache-free*... / de 40 à 45 pour cent des malades *n'ont* pour ainsi dire *plus de problème*... (*Canadian Magazine*, May 1990).

- sugar free, cafeine free, etc. / sans sucre, sans caféine, etc.

- We have also arranged for *interest free* monthly billings... / De plus, vous pouvez acquitter vos primes mensuelles ...*sans intérêt ou frais de financement*... (prospectus, CAUT group automobile / homeowner program).

- The cuisine ... leaves dieters *guilt-free*... / La cuisine *ne laisse pas de remords* aux personnes au régime ...(*En Route*, November 1990, p. 12).

2. CRÉATIONS LEXICALES.

Philippe Barbaud, qui voit là une ressemblance entre 'les créoles et le français parlé au Québec' (p. 53), ce qui porte à croire qu'il ne s'agit pas d'une tendance géographiquement circonscrite, fait observer que les Canadiens français ont tendance à créer des lexèmes négatifs, à partir de différents modèles :

a) **un pas + adjectif** : il est courant de former certains substantifs sur le modèle de : 'un pas-bon' (Barbaud 1987: 53 et 155) : 'une pas bonne', 'un pas fin', 'une pas fine', 'une pas belle', 'un pas- d'allure' (on peut même dire 'un vrai pas d'allure'), 'un pas intelligent', 'une pas montrable' (expression qui, soit dit en passant, n'existe pas à la forme masculine!', etc.

b) **un sans + nom** : 'un sans allure', 'un sans génie', 'un sans dessin' (= imbécile), etc. Il est intéressant de noter à ce sujet que, si ce type de création lexicale n'est pas nouveau, il semble s'être considérablement développé depuis la fin du siècle dernier (sans-abri, sans-parti, sans-patrie, sans-emploi, etc.)

3. LA NÉGATION PRÉFIXALE.

"Il [ce procédé] consiste à intercaler 'pas' suivi le plus souvent d'un adjectif 'psychologique'. La forme négative sert donc de PRÉFIXE passe-partout conformément à un usage extrêmement productif dans notre parler. C'est la raison pour laquelle j'utilise le trait d'union. Là aussi, l'énoncé suit la forme introductive. Cela donne : 'C'est pas-possible comme t'es beau!'; 'C'est pas-croyable comme tu ressembles à ton père!'; 'C'est pas-disable comme tu joues bien du piano!'; 'C'est pas-pensable comme on s'amuse à c'te place-là!', etc. [...] Dans un énoncé comme : 'C'est pas des farces comme le Canadien est bon cette année!', on retrouve le même procédé à la différence que le 'pas' est toujours associé au nom pluriel 'des farces'. Locution figée par excellence [...], celle-ci peut servir à exprimer un propos très positif autant que très négatif : 'C'est pas des farces comme il est

niaiseux!' Privé [sic] d'un énoncé à contenu émotif, l'expression 'C'est pas des farces!' véhicule habituellement un propos négatif." (Barbaud 1987:115)

4. REMPLACEMENT DU PRÉFIXE PAR LA NÉGATION PLEINE

Ici encore, Barbaud note qu'"il est courant de substituer la négation pleine au préfixe privatif des adjectifs du type : 'incroyable / pas croyable', 'imbuvable / pas buvable', 'impossible / pas possible', 'impensable / pas pensable' et surtout notre fameux 'pas capab'' qui supplante inévitablement 'incapable'." (Barbaud 1987:53). Cette tendance est peut-être plus accentuée au Canada que dans d'autres régions francophones, dans la mesure où elle peut donner lieu à des énoncés tels que : 'la viande n'est pas coupable', 'il est pas parlable' ou 'pas montrable', mais elle n'est pas circonscrite au continent nord-américain, comme M. Barbaud semble le penser. Elle se généralise, semble-t-il, dans tout le monde francophone: 'Ce n'est pas mangeable' semble avoir plus ou moins définitivement supplanté 'c'est immangeable'; 'ce n'est pas intéressant' ou 'c'est sans intérêt', 'inintéressant', etc. Et que dire des expressions si fréquemment utilisées telles que: 'c'est pas croyable!' (qui apparaît souvent au Canada sous la forme : 'ça se peut pas!'), 'c'est pas facile' (ou, sous une forme plus familière : 'C'est pas du gâteau / de la tarte', 'c'est pas un cadeau / de la petite bière / de la roupie de sansonnet', sans oublier naturellement le 'C'est pas pire' ou 'ça va pas pire', typiquement canadien.

5. LA DOUBLE NÉGATION

L'usage de la double négation est extrêmement répandu à l'oral chez les francophones d'Amérique du Nord, et semble en voie d'expansion, à en juger du moins par ce qu'en dit Philippe Barbaud : '**JE L'AI PAS DIT A PERSONNE**. "Qui donc oserait soutenir qu'avec un tel titre, j'abuse le lecteur si je prétends qu'il s'agit d'un énoncé caractéristique du parler d'ici? Sur le même modèle, à savoir l'absence de 'ne' et la présence simultanée de 'pas' et 'personne', il me serait loisible de citer nombre d'exemples attestant que cette forme de négation constitue la NORME orale du canadien-français. [...] D'autres exemples peuvent prêter main-forte à ma tournure initiale : 'J'ai pas rien vu'; 'Personne a pas pu rien nous dire'; 'J'avais pas aucune affaire de prête'; 'Rien est pas plus pire que ça', etc.

Tout le monde ici, à quelque couche socio-économique qu'il appartienne, fait un usage courant de cette tournure au point qu'elle en est banale.

Bref, voilà un usage de l'oral promu chez nous au rang de pratiques linguistiques non seulement familières mais soutenues. Il n'est pas rare en effet qu'il se glisse dans le français écrit de nos journaux ou dans les discours officiels de nos hommes politiques et de nos universitaires. C'est là un empiètement fort révélateur. S'il est vrai que le même usage se retrouve en France même, il reste néanmoins cantonné aux pratiques dites 'populaires'. Pas ici." (Barbaud 1987: 151-153).

Toutes ces créations lexicales, tous ces changements linguistiques tendent à montrer que les tournures négatives sont de plus en plus privilégiées par les francophones.

HYPOTHESES DIVERSES

RAISONS SOCIO-HISTORIQUES

Si la courbe de la négation en français monte en flèche au 20^{e} siècle, elle commence son trajet ascendant avec chutes à intervalles divers au 19^{e},et il nous paraît intéressant de

lier ce phénomène au sort réservé à la liberté de presse par les gouvernements successifs de cette époque. La modulation de la pensée par le biais de l'approche négative n'est-elle pas une excellente façon d'éviter la présentation directe de faits délicats, et d'éviter par là-même amendes, poursuites ou l'impossibilité de publier?

Sous l'Empire, Napoléon n'hésite pas à supprimer la liberté de presse: "Il ne subsista à peu près rien de la liberté de la presse. Nul article ne pouvait être publié sans être soumis à un censeur; nul ne put être imprimeur sans un brevet ou autorisation officielle." (Malet-Isaac, p. 406).

Sous Louis XVIII La Charte (2 mai 1814) rétablissait la liberté de la presse "sous la seule réserve de se conformer aux lois qui réprimeraient les abus de cette liberté." (id., p. 430). Dès 1815 cependant "la Chambre introuvable vota plusieurs lois d'exception, telle la loi sur les *écrits et les cris séditieux* (nov. 1815)." (id, p. 433).

En 1830, Charles X promulgue les Ordonnances de Juillet. "La première suspendait la liberté de presse et rétablissait l'autorisation préalable." (id, p. 437). La Charte révisée jurée le 9 août par Louis-Philippe contenait promesse de libéralisation sur la presse (id, p. 438). Mais par les lois de septembre 1835, les Français sont à nouveau soumis à des restrictions: "La censure et l'autorisation préalables furent rétablies pour les pièces de théâtre, les dessins et les caricatures. Ruinés par les procès et les amendes, les journaux républicains disparurent rapidement." (id, p. 443)

La 2e République rétablit en 1848 'l'entière liberté de la presse et de réunion' (id., p. 475). Mais en 1852 par décret Napoléon III "rétablit l'autorisation préalable, le cautionnement et la juridiction correctionnelle des délits de presse. Il institua un nouveau système de *répression administrative*: les journaux purent être avertis, suspendus, supprimés par simple décision des préfets." (id., p. 486). Enfin sous la 3e République "la liberté de presse fut assurée par la loi de 1881. La censure et l'autorisation préalable furent supprimées." (id.)

Ces dispositions libérales ont été cependant modifiées à deux reprises dans un sens restrictif (id., p. 545).

On aborde donc le 20e siècle en France avec encore des restrictions dans ce domaine et une aspiration constamment repoussée pour une liberté d'expression totale.

Pendant ce temps l'Angleterre jouissait d'une bien plus grande liberté d'expression. De plus la France est soumise à un système qui privilégie le bien général au détriment de la liberté individuelle, alors que les pays anglo-saxons grâce au *Bill of Rights* protègent les droits de l'individu. On peut voir donc à la fois dans le système anglais plus libéral et dans cette traditionnelle protection de l'individu dans le monde anglo-saxon, une des raisons possibles de la non-négativité de son discours par rapport au français.

Pour tenter d'expliquer l'énorme poussée de la négation au 20e siècle, il faut aller plus loin dans l'analyse et faire appel à d'autres facteurs. Certes, un siècle de 'modulation négative' forcée avait déjà modifié l'orientation de la pensée, mais il faut ajouter à cela la philosophie nihiliste et l'influence de la pensée Nietzschienne, et surtout Sartrienne sur l'intelligentsia et la pensée française. La critique de la connsaissance et du rationnalisme, la négation de Dieu, le rejet de l'autre, cette déconstruction générale de l'univers à la mode impliquent un discours chargé de négativité. Derrida, Adorno ... ont parlé de 'théologie négative' d'esthétique négative. Le recueil d'essais publié sous le titre *Languages of the*

Unsayable: The Play of Negativity in Literature and Literary Theory (1990) montre à quel point ce phénomène est complexe et profond.

Il est peut-être légitime d'avancer que ces mouvements de pensée ont beaucoup moins influencé la pensée anglo-saxone et de là son discours.

RAISONS CULTURELLES

Les Français sont très fiers de leur patrimoine culturel et linguistique, et la politique française de 'dédialectisation' des provinces a encore contribué à renforcer cette attitude de primauté du français. Les écoles ont gardé la tradition d'enseigner à bien écrire. Parmi les règles que tout élève devait respecter, on comptait entre autres, la non-répétion de termes ou expressions déjà utilisés et l'importance de lier les idées et d'introduire par une charnière logique toute nouvelle nuance de la pensée. Cette recherche d'équivalents et de liens constitue le problème le plus difficile à surmonter pour un élève. Il n'est pas étonnant qu'une fois de plus la modulation négative apparaisse comme une source importante d'alternatives. On apprend donc très jeune non seulement à utiliser ces formes, mais aussi à les valoriser. Le discours français est émaillé de nombreux 'il n'en demeure pas moins que', 'il n'est pas moins vrai que' ... Et il semble à la limite qu'un journaliste ou un écrivain ne soit pas satisfait de son texte s'il n'a pas utilisé quelques-unes de ces expressions.

RAISONS LINGUISTIQUES

Nous avons émis ailleurs[5] l'hypothèse que la fréquence des formes interro-négatives en français pouvait s'expliquer par le système d'opposition à deux termes pour l'affirmation. Par ailleurs, nous héritons d'une langue déjà formée et à travers les textes écrits de notre patrimoine et l'enseignement reçu, notre pensée se modèle. Plus la langue apprise est chargée de termes négatifs ou de tendances à l'approche négative et plus celui qui la parle en fait usage. Il semble qu'au départ dans ce processus, l'influence socio-politique ait joué un grand rôle, mais qu'une fois en marche, il se soit nourri de lui-même. Éducation, héritage littéraire et culturel, habitudes linguistiques acquises se sont renforcés mutuellement, comme un effet de boule de neige. Le courant historique qui a permis une plus grande liberté d'expression n'a pas eu pour effet d'arrêter ce processus, car il est arrivé à un moment d'éclatement où il ne suffisait pas de tout dire, il convenait au contraire de tout réfuter. Il a donc agi à nouveau comme un stimulant.

CONCLUSION

Cette étude a montré par des données chiffrées et des exemples précis à quel point l'ensemble du discours français était marqué par la négation, alors que les textes anglais l'étaient bien moins. Quelques hypothèses ont été avancées pour tenter d'expliquer ce phénomène. Il est certain qu'il est complexe et dû à de multiples forces qui s'unissent. On ne saurait donc l'expliquer dans sa totalité en quelques paragraphes. Cependant il a paru justifiable de lier ce phénomène à des raisons socio-historiques d'une part, en particulier l'absence de liberté d'expression au 19^{e} siècle en France ainsi qu'à l'éducation rigide des écoliers français en matière de stylistique.

Au 20^{e} siècle le besoin de tout remettre en doute et de tout contester ou nier qui a marqué la pensée française, est venu grossir une tendance déjà marquée à la négativisation. Tous ces mouvements se sont conjugués pour amplifier encore cet usage.

NOTES

1.

ligne 1 : fréquences absolues
ligne 2 : écarts réduits

	FREQUENCE / TENDANCE	PROSE			P.POETIQUE		VERS			TECHNIQUE			MONOLOGUE		DIALOGUE		RESTE	
		1800	1825	1835	1845	1855	1865	1875	1885	1900	1913	1922	1930	1935	1942	1955		
aucun	19366				15628		195		381		3162			3688		2945		12733
					-8.69		0.08		-12.44		17.01			1.01		-10.81		7.83
	+0.06	2255	1899	1377	876	967	913	754	1034	1376	1083	1337	1161	1160	1593	1581		
	-0.17	16.6	13.8	-0.3	-7.6	-5.8	-8.5	-11.0	-8.7	-0.3	-2.4	0.2	0.9	-0.7	7.7	2.1		
aucune	17335				13744		107		142		3342			3229		2493		11613
					-13.19		-5.07		-19.84		27.92			-0.44		-13.05		10.79
	+0.15	1988	1666	1118	808	989	845	750	980	1225	932	1190	1064	1058	1328	1394		
	-0.21	14.9	12.1	-3.7	-6.4	-1.6	-7.1	-7.9	-6.6	-0.5	-3.5	0.0	1.7	-0.1	4.3	1.4		
aucunement	531				378		4		4		145			61		74		396
					-7.28		-0.57		-3.55		10.51			-4.29		-2.55		5.51
	+0.64	10	19	18	15	50	26	55	13	28	28	36	42	28	40	123		
	+0.58	-5.3	-3.2	-3.3	-2.9	3.3	-1.2	4.5	-4.0	-1.7	-0.7	0.0	2.0	-0.8	0.6	13.2		
jamais	71645				61797		886		2584		6378			16141		14596		[illegible]
					22.90		6.33		-0.51		-27.76			25.85		15.02		-32.90
	+0.46	5271	5465	5431	3610	4487	3656	3787	4758	4574	4630	5607	4322	4803	5603	5641		
	+0.49	-9.4	4.1	4.1	-8.8	3.0	-12.0	-5.2	-3.1	-8.2	5.0	10.2	2.3	6.4	10.6	1.2		
ne	1033896				875268		11392		25586		121650			211672		213194		609030
					43.71		10.30		-63.40		-16.83			44.67		63.59		-86.92
	+0.59	8264	5760	7173	7405	3557	5979	5581	7855	0876	3947	7758	1681	8176	5546	1281	6640	8789	2181950
	+0.59	-12.5	5.0	-1.8	-27.3	-9.0	-23.7	-17.9	-30.0	12.7	24.7	22.0	4.2	12.6	32.8	6.6		
ni	56617				45028		811		1746		9032			10800		9942		35875
					-22.27		10.31		-7.10		26.38			1.93		-3.98		1.63
	-0.44	5105	4790	5527	3534	3464	3677	2606	2895	3779	2940	3470	2634	3415	3836	4945		
	-0.43	5.8	11.3	23.8	4.4	1.2	2.9	-11.6	-17.0	-4.6	-8.2	-6.8	-11.9	-0.9	-0.4	8.7		
non	75941				63223		624		2549		9545			13540		19183		43218
					1.53		-4.96		-4.21		2.17			-6.55		50.36		-34.95
	+0.97	4127	4156	4681	3535	3788	4022	3966	4775	5311	4909	6245	6021	6127	6521	7757		
	+0.96	-28.9	-18.6	-10.8	-13.5	-11.6	-10.2	-6.1	-7.0	-1.9	5.2	14.9	24.6	22.3	19.3	25.3		
nul	6464				4788		107		713		856			1103		1265		4096
					-19.22		5.28		31.70		2.28			-3.48		2.83		0.55
	+0.00	489	653	510	276	414	399	255	279	559	455	478	350	444	380	523		
	+0.00	-2.2	8.9	2.2	-5.2	1.3	0.0	-6.2	-8.2	4.5	3.4	1.7	-1.4	2.4	-3.0	1.0		
nulle	4358				3327		60		191		780			905		660		2793
					-11.78		2.49		2.60		11.22			3.39		-5.23		1.44
	-0.37	376	447	374	248	263	284	193	162	365	274	298	240	195	264	375		
	-0.34	0.6	7.7	3.5	-0.3	0.1	0.8	-3.7	-8.3	3.0	0.7	0.0	-0.9	-4.5	-1.9	2.1		
nullement	2555				2148		9		8		390			594		276		1685
					1.38		-3.29		-8.98		4.54			5.81		-9.69		3.05
	+0.31	113	225	178	85	247	162	120	139	118	154	267	140	225	183	199		
	+0.30	-7.1	3.0	-0.4	-5.3	7.8	0.3	-2.2	-2.9	-5.0	0.0	7.1	-0.7	5.6	0.6	0.0		
pas	508086				440986		5851		10937		50312			102034		114382		291670
					71.22		10.78		-56.70		-52.23			24.17		79.52		-83.09
	+0.82	2838	3302	6134	3452	4970	2793	1293	0129	0823	1727	3908	3756	5400	4634	0713	8114	4154	041663
	+0.89	-70.8	-35.0	-11.5	-27.2	-14.9	-12.5	-0.4	-19.4	14.1	41.3	28.9	26.6	40.9	38.3	11.9			
personne	18252				16488		137		249		1378			3843		3524		10885
					26.24		-3.40		-16.43		-19.57			7.94		3.85		-9.50
	+0.21	1052	1391	1328	1125	1015	1308	1018	1377	1205	1054	1442	988	1176	1270	1503		
	+0.38	-12.5	2.0	0.5	2.0	-2.4	5.4	-0.9	3.2	-3.0	-1.3	5.5	-2.4	1.7	0.7	2.4		
rien	95017				83830		935		2944		7308			21085		20344		53588
					42.57		-0.52		-8.96		-43.33			27.10		25.59		-42.36
	+0.87	4750	6406	6756	4822	5067	5627	5472	6550	7122	6902	6836	6168	7299	7553	7687		
	+0.81	-37.2	-5.8	-0.8	-9.6	-8.5	-3.4	0.2	-0.5	3.7	16.1	4.1	8.7	20.0	13.8	3.8		

TABLEAU RECONSTITUÉ À PARTIR DES DONNÉES RELEVÉES DANS UN TABLEAU GÉNÉRAL ALPHABÉTIQUE.

2. Le dictionnaire de Gunnel Engwall *Vocabulaire du roman français 1962-1968* donne le 11e rang à 'ne', le 17e à 'pas'. Dans la liste des adverbes les plus usités 'ne' et 'pas' arrivent en tête avec une fréquence absolue de 7658 et 4885 respectivement.

   A titre de comparaison le terme négatif de l'allemand 'nicht' arrive en 13e position, 'noch' en 33e, 'nur' le semi-négatif en 33e, d'après le dictionnaire du vocabulaire de la presse d'Inger Rosengren *Ein Frequenzwörterbuch der deutschen Zeitungsprache: Die Welt, Süddentsche Zeitung*, CWK Gleerup Lund Schweden, 1972. La place de la négation en allemand se rapproche, du point de vue de la fréquence, davantage du français que de l'anglais.

3. Le nombre de pages du magazine est supérieur à celui indiqué dans le tableau. Les pages de publicité, cartes et descriptions de restaurants qui apparaissent à la fin n'ont pas été dépouillés ni comptées. Il est évident que les nombreuses pages de publicité qui émaillent ces revues réduisent d'au moins un tiers le nombre réel de pages de texte.

4. C'est notamment aussi le cas de **safe** (sans danger, qui ne présente aucun risque, etc.). Les Canadiens français ont adopté un terme positif **sécuritaire**, mais qui n'a pas encore traversé l'Atlantique.

5 'Quelques aspects de la négation: comparaison de l'anglais et du français.' , à paraître dans *La Revue canadienne des langues vivantes*, 47, 2.

BIBLIOGRAPHIE

Barbaud, Philippe. 1987. *Le français sans façon*. Hurtubise: Montréal.

Brunet, Etienne. 1981. *Le Vocabulaire français de 1789 à nos jours*, d'après les donnés du TRÉSOR DE LA LANGUE FRANÇAISE. Slakine: Genève et Champion: Paris.

Budick, Sanford et Iser, Wolfgang, (éditeurs) 1990. *Languages of the Unsayable: The Play of Negativity in Literature and Literary Theory*. Columbia University Press: New York.

Engwall, Gunnel. 1984. *Vocabulaire du roman français (1962-1968): Dictionnaire des fréquences*. Institut de linguistique computationnelle: Université de Goteborg, Suède.

Johansson, Stig et Hofland, Knut. 1989. *Frequency Analysis of English Vocabulary and Grammar*, based on the LOB Corpus, Clarendon Press: Oxford.

Malet-Isaac, Béjean. 1950. *Histoire*. Hachette: Paris.

Rosengreen, Inger. 1972. *Ein Frequenzwörterbuch der deutschen Zeitungsprache: Die Welt Süddentsche Zeitung*, CWK Gleerup: Lund, Schweden.

WANDERING IN THE WORD-WOODS: SOME THOUGHTS ON THE ARRANGEMENT OF DICTIONARIES

Lewis J. Poteet
Concordia University, Montréal

ABSTRACT

Following my early experience of word-books, I chose the A-Z arrangement for entries in the small dictionaries I have written – The South Shore Phrase Books *(three),* The Far Eastern Townships Phrase Book, *and* The Hockey Phrase Book. *However, Connie Eble's* College Slang 101 *(1989) is arranged topically, using a form which led me to alter radically the format for* Parlons Hockey Talk, *the new attempt to record the hockey lexicon in both English and French.*

The hidden assumptions and unwritten messages behind the A-Z arrangement tend to perpetuate the popular assumptions that dictionaries claim a comprehensiveness which most lexicographers have renounced. Using alternative formats harmonizes with the primacy of vowel sounds in making linguistic atlases as well as acknowledging the concept and utility of 'keywords' in speech communities. Dictionaries and word-books need to acknowledge more explicitly their function as guides to the pathways among language units in a larger linguistic reality than they may ever map completely.

When I was young, I spend a lot of time alone, wandering and reflecting. First I wandered in the Bible and, as it enjoined, meditated on it. Later, I wandered mostly in libraries, and reflected on many things. I recall during what seemed a time of flirtation with madness, thinking that I would know that I had gone mad if I ever found myself reading the dictionary from the beginning, as if it were a novel. At the same time, I perceived that someone might well choose to do that very thing.

So it has come as a slight shock to me to consider that I now do in fact read some dictionaries from the beginning to the end, making me, by my own early rule, certifiable. Worse yet, I write dictionaries, or rather 'phrase books'; the difference is, as they say,'academic', or better yet, say, 'popular', not necessarily the same thing. Of course the dictionaries I mean – the ones I read through and write, are small, specialized, regional or otherwise limited samplings of the lexicons of speech communities, whatever they are, little Towers of Babble, or games of partly broken telephone.

When I first drafted the manuscript of the *South Shore Phrase Book*, around 1978 or 1979, the format that I used was the one I considered self-evident and universal for all word books: it was an alphabetical arrangement of the entries. As if to emphasize the rightness of my choice (but in fact unconsciously, because it just seemed appropriate, I prefaced the As ('able', 'aboard' etc.) with a folk rhyme, 'A apple pie, B bit it, C cut it' and so forth... which pointed to the scheme in imagery both universal and appetizing. So 'A' to 'Z' was my books' plot, every one of the South Shore books, and thus by default also the *Hockey Phrase Book* I did with my son Aaron and published in 1987, and the manuscript of car slang I am doing with my brother Jim.

I was not unaware that other formats were possible. But it was clear I did not like them. For some reason I early conceived a distaste for Roget's Thesaurus, that beloved recourse of many students and writers, and I think it was that it interrupted the alphabetical order I preferred.

In fact, it was not until I was teaching a graduate course on lexicography and listening to a guest lecture by Eyvind Ronquist, the medievalist and word-wizard in my department, a lecture on early word-books, that I was reminded of some of the formats once used and almost universally discarded. As you no doubt know, they were random-seeming lists of things, from angels to rocks; cookery terms; and so forth. Their lack of mutual exclusiveness, their overlappings, the illogicalities of their classification schemes – all made me think less of them than my old familiar A-Z, and even at this time when I may have broadened my perspective I had no thought of ever abandoning my traditional ways of laying out my work.

No, it was not until I saw a radically contemporary work – for me even more vangardist than Cheris Kramarae's *Feminist Dictionary* (which for all its breaking new ground, was arranged traditionally, A-Z!) – it was Connie Eble's *College Slang 101*, which Spectacle Press published in 1989, that made me discover I wanted to try something different. This stimulating little book happened to appear at the same time as the other 'first' collection of college slang, *UCLA Slang*, which was arranged A-Z, making comparisons easy. They are comparable in another interesting way: *CS 101* was collected in North Carolina, whereas *UCLA Slang* comes, of course, from the fountainhead (or heads) of young peoples' talk, the spot on the planet where Valley Girl met the Silver Surfer, and their younger brothers and sisters of the Skate or Die persuasion spread the rap through every mall, video arcade, and campus on the continent north of the Rio Grande. Yet, curiously, I find *CS 101* the more interesting of the two, and in a way more accessible. After, that is, I had made an index for it, an alphabetical index. I set out to read through both of them, but I didn't put down *CS 101* until I had finished it, whereas *UCLA Slang* is close to becoming a 'reference work' in my library, a term which indicates that it falls short of the gold ring in interest, if not usefulness.

I must now give you an idea of the arrangement of *CS 101*: in three major 'entry' chapters, that is, chapters made up almost entirely of entries, and titled 'The Origins of College Slang', 'The Meanings of College Slang', and 'Why College Students Use Slang', Eble illustrates her account of the nature of the speech community from which she collected during the 1970s and 80s. Some of the categories are based on form – words that add '-topicalage' like 'foodage', or start with 'mega-'; words that are formed by combining them with other words, like 'action' – e.g., 'food action', or 'doy-burger' (a dimwit or physically uncoordinated person), or 'earth daddy' (older-than-college-age male who displays a 60s style of mellowness, wears a beard, and attends college parties), or '-out', as in 'goob out' (to cause repulsion or disgust). In subchapter titles that are vaguely and I think disquietingly reminiscent of those cutesy menus at America's interstate interchange restaurants, she makes the name perform the action described: 'Slangdom: Adding Prefixes and Suffixes', 'Slanguage: Putting pieces of words together', 'To the min: Shortening Words and Phrases', and 'Budwiper: Playing with sounds'. In the section on meaning, the categories are topical – drugs, drinking, studying, Greek Speak, eating, vomiting, hello and goodbye, and so forth. The last short section, 'Why College Students Use Slang', samples 'evaluative words, retorts, and words that express agreement or support'. Words appear in more than one section, as is appropriate. I found the book stimulating, and frustrating. I enjoyed testing it on my friends and students, and them on it; but inevitably someone would say, 'Has she got... megabitch', for example; and unless I remembered exactly, I wasn't sure, and I didn't know how to find it without scanning the whole book!

Eventually my frustration produced action: I sat down at my computer, and made a damn index myself. I worded with a furious intensity for four hours, and when it was done, I popped a copy into an envelope and set off up the street to mail it to Connie Eble... and nearly fell down from the leg cramps! But now that my copy has an index, I use it with perfect joyfulness.

It happens that at the moment *CS 101* appeared in my life, the publisher of the *Hockey Phrase Book* had just gone bottoms up, and I was searching for a new publisher. I found one, Nu-Age Books in Montréal, and one of the owner/editors, Karen Haughian, seeing *CS 101*, urged me to consider finding something like that for the new, as it was to be, bilingual, expanded version *Parlons Hockey Talk*. At that moment, distressed as I was at the aborted brilliant career of the *HPB* (it had sold 3000 copies in its one year of life in print), she looked like my fairy godmother, and I was inclined to hang on her every word. What's more, I was beginning to reflect for the first time on the hidden assumptions and unwritten messages behind the text, the form, of most dictionaries.

What the form says, not very loudly unless you are listening, is 'come wander in my word-world, my complete list, my alpha to omega. I have it all. You want to know? Look it up here.' The symmetrical order has the air of a military procession, a ritual of truth and authority. In fact, before I even meditated the social sacrilege of possibly reading a dictionary like a novel, I learned to read by reading, five times in a row, straight through from beginning to end, the Bible, the Christian book of history, truth, and all things. And this book of authority over, in my Nazarene faith, all things sacred and secular, was primarily taught me as an encyclopedia, like the dictionary, a complete guide to it all, to be consulted at need, a work of reference.

Yet all we learn as students of language gives the lie to this parade of authority. No word-book has all the words. No word-book ever has and none ever will. Even the A-Z implied equality to all letters is a fake: as Raoul Duguay so memorably said, the 'language of the vowels' is a way of talking to anyone, across language boundaries. (Duguay, who seemed to speak for a segment of Québécois youth, said, in the turbulent early seventies in a performance at Concordia University in Montréal, "I am not going to speak to you in English or French; I am going to speak to you in the language of the vowels.") He was of course introducing his sound poetry, but during the debates over the language laws in the late 80s, a CBC panellist claimed, "Québécois is just a way of pronouncing, really.", no doubt considering the many American words that enter the French language without passing customs at the border with Vermont, New Hampshire, and New York. There are fewer vowels than consonants, and the way we say them is principally how we tell where someone's from. I see that it is variation in vowel-pronunciation, mainly, that is most prominent in linguistic atlas work. So vowels, I am saying, are more important than other letters, in a way; though I love and have devoted my career to the lexicon, and have been reluctant to join the international Phonetic Battalions, I must say that words seem to be able to move freely between languages, but vowel sounds mark speakers as to origin, or the ability to mimic origin, in a clear and distinct way. All letters are equal, but some letters are more equal than others!

The same is true of words. Some words are 'keywords' in a group, a speech community or culture. As I collected words and phrases along the South Shore of Nova Scotia, I discovered that some words have more stories about them than others. Why are Nova Scotians called 'bluenosers'? At least three main accounts are widespread. It is less important to me which one is the 'true' account of origins than that there are more stories, there is more talk, about this word than about others. Westerners – see Chris Thain's *Cold As A Bay Street Banker's Heart: The Ultimate Prairie Phrase Book* – find 'outfit' as useful a word as 'haywire' was to the farmer, as a thing, a concept, and a word, too! Terms of

rough measure are keywords. In fact, considering the two billion speech events a day in English, in which words are shortened, combined, misused but in contexts which convey meaning accurately anyway, and so forth, every word that finally makes it into a dictionary is in a sense a keyword. All words are equal, but some words are more equal than others.

I am sure that we all agree that it is more honest of the lexicographer to give up the pretence of encyclopedic universal inclusion and to try to dislodge the popular notion that dictionaries claim it. But we love the idea of wandering in the woods, of a world with wilderness, of a book of words in which you may find anything. I read a dictionary like a novel, but I'm strange; I have to be tricked into wandering. And here I do notice that I love the individual voice we encounter in the small phrase book: it is a guide, a friendly voice which leads us ahead, and often does so by showing how one best-known, 'key' word leads to others. 'Botts Dots: see also cats' eyes'; 'Claquage: see also Charley-horse, and Donner une jambette'. A dictionary may offer a voice, a narrative, with many digressions, a true garden of forking paths, if not as I once believed or hoped, a circular library with everything in it that ever was, or is, or will be.

REFERENCES

Eble, Connie. 1989. *College Slang 101*. Spectacle Lane Press.

Munro, Pamela (ed.) 1989. *U.C.L.A. Slang*. Department of Linguistics, University of California, Los Angeles Occasional Papers in Linguistics 8.

Poteet, Lewis J. and Aaron C. Poteet. 1987. *The Hockey Phrase Book*. Montréal: Eden Press.

Poteet, Lewis J. 1983, 1985, 1988. *The South Shore Phrase Book*. Hantsport, N.S.: Lancelot Press.

Thain, Chris. 1989. *Cold as a Bay Street Banker's Heart: The Ultimate Prairie Phrase Book*. Saskatoon: Western Producer Prairie Books. 1987.

PARLONS HOCKEY TALK BOOK PLAN

1. **What it is/Le hockey, c'est quoi?** *Hockey, bandy, shinny, old time hockey.*

2. **Who? People/Qui? Gens, sobriquets, positions, rôles, lignes.** *Nicknames. Roles, jobs, positions. Lines.*

3. **Where? Places/Où?** *Homerdome, gondola, catwalk, CH, alleys, in the cell.*

4. **When? Then and now: hockey history in talk/Quand? Le hockey d'antan et aujourd'hui.** *WHA. Old time hockey. Team Surfboard.*

5. **What is that? Things/C'est quoi, là?** *Zamboni. Jill. The biscuit. Hockey cushion. Iron lung. The five hole.*

6. **Tricks/Truc et jeu**

   a. **Showing off/S'exhiber.** *Showboater. Stickhandling. Deke. Finesse players. Taking a dive.*

   b. **Dirty tricks/Jeux sales.** *Coldcocking. Blindsiding. Highsticking. Dr. Hook. Give someone a Brad. Donner une jambette.*

7. **Wrong moves/En erreur.** *Pull a Zamboni. Scheiss-eie. Sieve. No Smoking Section. Take the gate.*

8. **Ouch! – injuries/Les bo-bos.** *On the limp. Claquage. Pitchfork. Harponnage.*

9. **How to. Advice, rules, and lore/Comment? Proverbes, regles, conseils.**

10. **Have another donut! – insults/Les insultes.**

11. **Technical terms with a difference/Le lexique technique mais pas scientifique.** *Breakout. Dipsy-doodle. Get and go. Follow-in. Forecheck in tandem. Pick-play.*

12. **Hockey talking: words about talk/Le parler hockey.** *Chirping. Bas. Elmer. Exchanging pleasantries.*

LA DÉNASALISATION DES VOYELLES /ɛ̃/, /ɑ̃/ ET /ɔ̃/ OBSERVÉE DANS UN PARLER ACADIEN DE L'ÎLE-DU-PRINCE-ÉDOUARD

Robert Ryan
Université de Waterloo

Barry Miller
Université York

RÉSUMÉ

Parmi les particularités que peuvent présenter en acadien les réalisations des voyelles nasales /ɛ̃/, /ɑ̃/ et /ɔ̃/ figure une tendance à la dénasalisation dans certains contextes précis, notamment en syllabe ouverte inaccentuée. La présente communication entreprendra d'identifier les différentes formes dénasalisées de ces trois voyelles, de même que l'ensemble des contextes sonores où elles se manifestent chez six locuteurs acadiens de la Région Évangéline de l'Île-du-Prince-Édouard, représentant les deux sexes et trois tranches d'âge différentes. L'on s'efforcera, par la même occasion, de mesurer la stabilité ou la variabilité de la dénasalisation des trois voyelles au sein de ce groupe d'informateurs, en cherchant à mettre ces phénomènes en rapport avec les différences de sexe et d'âge représentées.

1. INTRODUCTION

Les particularités phonétiques et phonologiques que présentent les trois voyelles nasales /ɛ̃/, /ɑ̃/ et /ɔ̃/ des parlers acadiens des quatre provinces atlantiques du Canada ont déjà fait couler beaucoup d'encre. Quiconque examine le nombre non-négligeable d'études consacrées à la description et au fonctionnement de ces trois voyelles est frappé tout d'abord par le grand éventail des réalisations sonores qu'elles peuvent présenter en comparaison du français standard. Sur le plan phonologique, ce qui retient avant tout l'attention, c'est la variabilité des oppositions fonctionnelles entre ces trois voyelles, selon les différents contextes syllabiques et accentuels où elles peuvent figurer. En effet, en syllabe ouverte accentuée le système à trois voyelles nasales attesté dans les autres contextes sonores se réduit le plus souvent, dans les parlers acadiens les plus conservateurs décrits jusqu'ici, à un système à deux voyelles, à savoir /ɛ̃/ et /ɑ̃/. Là où elle se produit, cette réduction résulte le plus souvent de la neutralisation, dans ce contexte précis, de l'opposition entre les voyelles /ɔ̃/ et /ɑ̃/ à la faveur de /ɑ̃/.[1]

2. OBJECTIFS DE L'ÉTUDE

Dans cette communication nous nous proposons d'examiner une autre manifestation de la variabilité phonétique et phonologique que peut présenter le système des voyelles nasales de l'acadien, à savoir, une certaine tendance à la dénasalisation partielle ou totale observée dans les parlers acadiens de l'Île-du-Prince-Édouard avant tout en syllabe ouverte inaccentuée. Si la neutralisation déjà signalée en syllabe ouverte accentuée affecte les rapports entre les voyelles nasales, et notamment entre /ɔ̃/ et /ɑ̃/, la perte de nasalité que

nous allons décrire est susceptible d'affecter, elle, l'on s'en doute, les rapports entre les trois voyelles nasales et les voyelles orales correspondantes et, de ce fait, la rentabilité de l'opposition vocalique orale-nasale.

Nous nous efforcerons d'identifier dans le corpus acadien retenu la totalité des réalisations dénasalisées des voyelles qui seraient systématiquement nasales en contexte identique en français standard. Nous entreprendrons aussi d'établir une correspondance entre ces diverses réalisations dénasalisées et les différents contextes qui semblent en favoriser l'apparition. Nous essaierons enfin de mesurer la stabilité du phénomène à l'étude à travers trois générations de locuteurs représentant les deux sexes, mais appartenant essentiellement à la même classe socio-économique.

3. IDENTIFICATION DU TYPE DE DÉNASALISATION À L'ÉTUDE

Commençons par serrer de plus près le type de dénasalisation à l'étude, en bien le distinguant d'autres formes de dénasalisation déjà clairement identifiées en acadien et amplement décrites. Dans bon nombre des parlers acadiens analysés jusqu'ici les voyelles nasales /ɛ̃/, /ɑ̃/ et /õ/ peuvent se réaliser en syllabe ouverte accentuée sous forme d'une voyelle orale ou mi-orale, le plus souvent, mais pas forcément apparentée à la voyelle nasale en question, suivie d'une des consonnes nasales [m], [n] ou [ŋ]. C'est ainsi que Karin Flikeid signale dans son excellent article intitulé 'Unity and Diversity in Acadian Phonology: An Overview Based on Comparisons Among the Nova Scotia Varieties.' (1989:90), entre autres, des prononciations comme [ɑ̃ŋ] pour /ɑ̃/ ou /õ/; [ɔŋ] pour /õ/, ou [ɔn] pour /ɛ̃/, cette dernière réalisation apparaissant sous l'accent dans l'alternance sonore si caractéristique de la Baie Sainte-Marie, en Nouvelle-Écosse, [mɛ̃] ~ [mɔn] 'main'; [pɛ̃ ~ pɔn] 'pain', etc. De même, à l'Île-du-Prince-Édouard nous avons relevé assez fréquemment des réalisations dénasalisées en syllabe accentuée comme [æ̃ŋ] dans [marʃjæ̃ŋ] 'marchions'; [krejæ̃ŋ] 'crayons', etc.[2] Notons, au passage, la ressemblance entre cette sorte de dénasalisation en acadien et celle qui caractérise le français méridional où, par exemple, 'bien' peut se réaliser [bjæ̃ŋ], 'marchand' [marʃɑ̃ŋ], etc.

La dénasalisation qui fait l'objet de notre étude présente au moins trois caractéristiques qui la distinguent nettement de celle que nous venons de décrire. Tout d'abord, phonétiquement elle se caractérise par une perte partielle ou totale de son timbre nasal, sans cependant aucun ajout de consonne nasale. Deuxièmement, la voyelle mi-orale ou orale qui en résulte conserve une articulation non-diphtonguée très proche de celle de la voyelle nasale correspondante. Enfin, à la différence de [æ̃ŋ] ou [ɔŋ], etc., la dénasalisation en question se manifeste le plus souvent en syllabe *inaccentuée* ouverte.

Remarquons que ce type de dénasalisation figure peu ou pas dans les analyses effectuées par d'autres acadianistes. Il n'en est nullement question, par exemple, dans la synthèse minutieuse qu'a faite Karen Flikeid (1988) des systèmes sonores des parlers acadiens néo-écossais. Francis Landry (1985:103) n'en signale, pour sa part, qu'une seule occurrence, à savoir, [kɔbɑ̃j] 'combien' parmi les centaines d'exemples de voyelles nasales sur lesquels repose son étude. Vincent Lucci (1973:65), quant à lui, ne semble pas avoir observé ce phénomène dans la région de Moncton, du moins en syllabe inaccentuée. Il constate, cependant, une tendance à la dénasalisation de /ɛ̃/ en [ɛ̃] en finale libre dans des mots comme [matɛ̃] 'matin', [desɛ̃] 'dessin', et [rɔʒɛ̃] 'regain'.

4. PARLERS DE L'ÎLE-DU-PRINCE-ÉDOUARD, INFORMATEURS ET CORPUS

Avant d'examiner des exemples de cette dénasalisation et d'analyser plus en détail ce phénomène acoustique et articulatoire, situons brièvement les parlers acadiens de l'Île-du-Prince-Édouard, présentons les six informateurs et caractérisons le corpus que nous ont livré ces derniers.

L'Île-du-Prince-Édouard comptait, lors du dernier recensement complet en 1981, environ 6 000 habitants de langue maternelle française qui représentaient environ 5% de la population de l'île.[3] Cependant, le fait que seulement un peu plus de la moitié de ces locuteurs déclaraient utiliser le français régulièrement à la maison fournit un indice inquiétant de la rapidité avec laquelle ces francophones s'assimilent à la majorité anglophone.

Ce sont dans la région dite Évangéline et dans celle de la petite ville de Tignish situées dans la partie nord-ouest de l'île ou le Comté de Prince que sont établis la très grande majorité des Acadiens. De ces deux régions c'est de toute évidence celle d'Évangéline où la langue et la culture franco-acadiennes semblent les plus vivaces. Or les six informateurs dont les témoignages constituent le corpus oral d'environ 60 minutes sur lequel s'appuie cette étude habitent tous la petite localité d'Abram-Village qui se trouve au coeur de la Région Évangéline. Ces témoignages, de longueur égale et de caractère très spontané et naturel, ont été recueillis par deux enquêteuses acadiennes, natives de la Région Évangéline, dans le cadre d'une enquête sociolinguistique de grande envergure portant essentiellement sur les deux régions acadiennes de l'île et impliquant une cinquantaine d'informateurs.[4] Parmi les 24 informateurs d'Abram-Village, les six sur lesquels porte cette étude ont été choisis de manière à représenter les deux sexes au sein de trois tranches d'âge, le premier groupe ayant moins de 30 ans, le deuxième ayant entre 30 et 59 ans et le troisième plus de 60 ans. Tous les informateurs retenus appartiennent *grosso modo* à la même classe socio-économique. Les informateurs de la génération la moins âgée ont terminé leurs études secondaires dans une école francophone d'Abram-Village. La génération du milieu a une scolarité moyenne de 10 ans, tandis que, dans le cas des locuteurs les plus âgés, cette scolarité est réduite à 6 ans.

5. ANALYSE DES DONNÉES

Notre corpus a livré au total 1187 occurrences de voyelles qui seraient normalement en français les nasales /ɛ̃/, /ɑ̃/ et /õ/. Parmi ces occurrences, nous en avons relevé 157, soit 13,2%, qui accusent une perte de nasalité, cette dénasalisation ayant été décelée à l'oreille. Il y en a 64 où la perte de résonance nasale n'est que partielle, les 93 autres étant, de toute évidence, complètement orales. Voici quelques exemples en contexte de voyelles dénasalisées observés chez l'informateur de 42 ans, l'informatrice de 72 ans et l'informateur âgé de 77 ans.

[o klɛr la fɑt^{h}ɛn sɑ setɛ ... ɑ̨ ɑ̨ ʃɑ̃ntę̨ sɑ̨]

[dɑ̃̃ le ... dɑ̃̃ le le barʒ də pɑ:j]

[truvjõ /i vənjã nu ʃarʃe /i nu truvjɑ́]

[bɛ̨ ʒ kwɑ k wi /bɛ̨ ʒ kwɑ]

La dénasalisation attestée n'affecte pas les trois voyelles /õ/, /ɑ̃/ et /ɛ̃/ dans la même mesure. C'est la nasale /õ/ qui semble être de loin la plus sensible à la perte de nasalité, cette modification de timbre affectant 63 ou légèrement plus de 15% de ses 416 occurrences chez les six locuteurs, environ 45% de ces réalisations étant, du reste, complètement dénasalisées ou orales.

Vient en deuxième position la voyelle /ɛ̃/, dont un peu plus de 13% des 255 occurrences voient leur nasalité réduite ou supprimée. Ce pourcentage comprend 39 occurrences de l'interjection [bɛ] → [bɛ̃] 'bien' en français standard, dont la dénasalisation est systématique au point de faire de [bɛ] ou [be] une unité à voyelle orale très stable. Si l'on fait abstraction des occurrences de [be/ɛ], la fréquence de /ɛ̃/ dénasalisé diminue à environ 6%, dont à peu près la moitié (47%) des occurrences perdent complètement leur nasalité.

Si les 516 occurrences de ce qui serait en français standard la voyelle nasale /ɑ̃/ font de celle-ci la nasale la plus fréquente des trois dans notre corpus, sa tendance à la dénasalisation est toutefois très restreinte, car elle ne touche que 40, soit 0,8% de ses occurrences, celles-ci étant complètement dénasalisées à 62,5%.

6. CONTEXTES ENTRAÎNANT LA DÉNASALISATION

Les données de notre corpus révèlent une corrélation étroite entre la dénasalisation des trois nasales /õ/, /ɛ̃/ et /ɑ̃/ et leur emploi en syllabe ouverte inaccentuée. En effet, 71% des occurrences dénasalisées figurent dans ce contexte précis. C'est la syllabe ouverte accentuée qui constitue la deuxième en importance des contextes favorisant la dénasalisation. Il y a toutefois un écart très important entre les quelque 19% des cas de dénasalisation attestés dans ce contexte précis et les 71% observés en position inaccentuée. De toute façon, ce qui semble déterminant, c'est le fait que la nasale s'emploie en syllabe ouverte.

7. RAPPORTS ENTRE LA DÉNASALISATION ET L'ÂGE

La mise en rapport du phénomène de la dénasalisation avec l'âge des six locuteurs retenus révèle chez les trois hommes une grande stabilité du phénomène à travers les trois générations examinées, les pourcentages étant, à commencer par la génération la plus âgée, 15,5%, 23,2% et 23,3%. En revanche, chez les trois locutrices, si la plus âgée a enregistré un taux de dénasalisation de 16,4%, donc légèrement plus élevé que celui de l'homme de la même génération, la tendance à la dénasalisation accuse une diminution sensible et progressive chez les deux autres locutrices, à savoir, 9,6% chez celle de 35 ans et 3,2% chez la plus jeune.

Si cette réduction progressive de la dénasalisation chez les trois locutrices va à l'encontre de la stabilité du phénomène observée chez les trois hommes, elle correspond toutefois à la tendance attestée chez deux garçons, représentant la plus jeune des quatre générations d'une famille acadienne d'Abram-Village, dont nous avons déjà étudié le système des voyelles nasales.[5] Par conséquent, en l'absence d'autres facteurs sociolinguistiques évidents, l'on peut se demander si cette diminution de la dénasalisation chez les locutrices n'est pas attribuable à une certaine influence normalisante du système scolaire. Une analyse de la totalité du corpus d'Abram-Village devra permettre d'éclaircir

cette question et de mettre peut-être en évidence d'autres facteurs sociolinguistiques que ce corpus limité ne permet pas de cerner.

8. CONSÉQUENCES PHONOLOGIQUES

Avant de conclure, examinons brièvement les conséquences éventuelles de cette tendance à la dénasalisation pour le fonctionnement phonologique du système des voyelles nasales en syllabe ouverte et avant tout inaccentuée.

Là où la dénasalisation est considérable ou totale, l'on peut observer une neutralisation de l'opposition entre chacune des trois voyelles nasales et la voyelle orale correspondante, c'est-à-dire, entre /ɛ̃/ et /ɛ/, entre /õ/ et /ɔ/ et entre /ɑ̃/ et /a/ ou /ɑ/. Il en résulte donc une certaine instabilité du fonctionnement de ces nasales dans ces contextes. Nous avons cependant relevé jusqu'ici très peu de cas où le contexte lexical et la structure morpho-syntaxique ne permettent pas de dissiper toute ambiguïté. Citons, à cet égard, des formes verbales comme 'entendre' et 'entendu' que notre informateur de 77 ans a réalisées [atɑ̃d] et [atɑ̃dɛy], c'est-à-dire, phonétiquement identiques à 'attendre' et 'attendu'. Cependant, celles-ci ne sont pas ambiguës dans le parler d'Abram-Village dans la mesure où elles y sont réalisées [ɛspare] 'espérer', 'espéré'.

9. CONCLUSION

L'enquête que nous avons menée jusqu'ici permet donc de conclure que cette tendance à la dénasalisation constitue une caractéristique non-négligeable du système des voyelles nasales de l'acadien d'Abram-Village, surtout en syllabe ouverte inaccentuée. Du reste, une analyse spectrographique devra permettre de déterminer encore plus précisément le degré de dénasalisation qui affecte ces voyelles. Répétons enfin que, malgré la neutralisation qui en résulte, cette tendance à la dénasalisation ne semble avoir qu'une incidence négative très minime sur la transmission efficace du message.

NOTES

1 Karin Flikeid (1988:88) fait remarquer, cependant, qu'à Petit de Grat, dans l'Île-du-Cap-Breton, en Nouvelle-Ecosse, ce sont les voyelles /ɑ̃/et /ɛ̃/qui se neutralisent en opposition à /õ/, tandis qu'à Chéticamp ce sont les trois voyelles qui se neutralisent dans ce contexte précis.

2 Se reporter à King et Ryan (1986:101 - 102).

3 Se reporter à Statistique Canada: *Recensement du Canada* 1981, Document E-572.

4 Cette étude, en cours depuis plus de trois ans, est subventionnée par le Conseil de Recherches en Sciences Humaines du Canada (Subvention # 410-87-0586).

5 Voir Ryan (1989:89-102).

REFERENCES

Flikeid, Karin. 1988. 'Unity and Diversity in Acadian Phonology: An Overview Based on Comparisons Among the Nova Scotia Varieties '. *Revue de l'Association de linguistique des provinces atlantiques (RALPA /JAPLA)* 10: 64-110.

Gesner, B.Edward. 1986. *Bibliographie annotée de linguistique acadienne.* Québec: Centre international de recherche sur le bilinguisme.

King, Ruth et Robert Ryan. 1986. La phonologie des parlers acadiens de l'Île-du-Prince-Édouard . *Actes du dixième colloque annuel de l'Association de linguistique des provinces atlantiques (ACALPA /PAMAPLA)* 10: 95-108.

King, Ruth et Robert Ryan. 1987. 'The construction of a rural sociolinguistic corpus: The Prince Edward Island Study.' *Methods in Dialectology* (Alan R. Thomas, ed.). Clevedon /Philadelphia: Multilingual Matters 95-108.

Landry, Francis. 1985. *Étude synchronique des voyelles nasales dans le parler de Pubinico-Ouest.* Thèse de maîtrise inédite. Université de Montréal.

Lucci, Vincent. 1973. *Phonologie de l'acadien (parler de la région de Moncton, N.B., Canada).* Paris: Didier.

Ryan, Robert. 1989. 'Stabilité et évolution phonologiques observées dans le système des voyelles nasales de quatre générations d'une famille acadienne de l'Île-du-Prince-Édouard (Canada)'. *Actes de langue française et de linguistique* 2:90-102. Halifax: Université Dalhousie.

Statistique Canada. 1981. *Recensement du Canada.* Document E-572. Ottawa.

Séguinot, André. 1968. 'Etude sur le degré de nasalité des voyelles nasales en français canadien et en français standard'. *Recherches sur la structure phonique du français canadien* (Pierre R. Léon, rédacteur). Montréal /Paris /Bruxelles: Didier.

COMPUTER GLOSSING AS A TOOL IN CONVERGENCE RESEARCH

Ian Smith
York University

ABSTRACT

This paper reports the use of computer glossing to assist in the generation of parallel texts for an ongoing research project on Dravidianization in Sourashtra (Tamil Nadu Gujarati).

The initial data base is a corpus of 35 hours of recorded conversations in Sourashtra. The machine glossing system inputs a Sourashtra text and, after performing some preliminary morphological analysis, generates a Tamil gloss on a simple unit-by-unit basis from a computerized glossary. The machine output is given to bilingual assistants as a basis for their translation. The approach is particularly suited to work employing linguistically untrained assistants.

1. INTRODUCTION

This paper reports on the use of a computer glossing system as a research aid to the study of convergence. The system described here was developed for use in an ongoing study of Tamil and Sourashtra convergence.

Sourashtra is a language spoken by migrants from the Saurashtra region of what is modern Gujarat, who according to tradition, left their homeland in the 11th century and after a series of halts in Marathi- and Telugu-speaking areas arrived in the Tamil-speaking region of Southern India following the defeat of the Vijayanagar Empire at Talikota in 1565. The largest group of speakers is found in Madurai, where they form 20% to 25% of the city's population of one million. Madurai is located about 450 kilometers southwest of Madras, and it is here that the research is being undertaken.

2. CONVERGENCE RESEARCH METHODOLOGY

The methodology of the study is similar to that generally employed in convergence research and involves two stages:

Stage I. The contrastive analysis stage: Sourashtra and Tamil are compared in order to determine the points of similarity of the two languages

Stage II. The historical analysis stage: earlier stages of Sourashtra are reconstructed to establish which similarities may be due to the transference of linguistic characteristics from Tamil (the possible earlier influence of Marathi and Telugu will also have to be examined).

Here we are concerned with only the first stage. The contrastive analysis may be carried out in two main ways. One is to work systematically through the structures of the

two languages, hoping nothing is missed. The other is to work from parallel texts in the two languages, looking for points of similarity and difference, again hoping that they contain examples of all the structures in the two languages. The best approach is probably a combination of the two techniques, but the exact weighting of each would depend on the investigator and the ultimate purpose of the research. For example, if the investigator knows both languages well, or has access to linguistically trained bilinguals, the first technique might yield the quickest results. Even so, the subsequent analysis of texts is important as a test of the validity of armchair analysis, and as a source of new structures and usages. Certainly, if the investigator does not know one of the languages well, and has no access to linguistically trained bilinguals, then a primarily text-based approach is likely to be more reliable. This is because linguistically untrained bilinguals are often unable to work with examples out of context, they are often unable to think of language in structural terms, and they may tend to simplify or distort the language for the supposed benefit of the investigator. On the other hand, with a text-based approach, provided the texts are natural, you know you are working with real language.

For the Sourashtra-Tamil convergence study the initial data is a corpus of approximately 35 hours of recorded conversations and insider interviews in Sourashtra, all involving naturally occurring groups. These have yielded approximately about 70,000 lines of typescript, or about 150,000 lines of bilingual text, when Tamil translations are added.

One of the uses of a corpus of bilingual texts is to gauge the extent to which the languages are comparable in actual practice, that is, the degree to which they can be intertranslated on a morph-by-morph basis.

With this goal in mind, we want our translations to be carefully controlled so that a given morph is translated consistently, to the extent that this is possible without distorting the meaning or violating the norms of usage.

Linguistically untrained bilinguals are in general unable to provide such consistent and accurate translations. For example, a group of five educated bilingual research assistants asked to provide Sourashtra equivalents for an extensive paradigm of Tamil verb forms often required considerable discussion to reach consensus on a particular form, and even then were far from completely accurate. Thus, a future perfective paradigm was given for a present perfective, a perfective imperative plural for a second person plural future perfective, a masculine form instead of a feminine past perfective singular, a first person plural future emphatic for a first person plural present, a perfect reflexive paradigm instead of a past reflexive, and so on. Similar errors appeared in translations of texts. The problem is that when more than one form can be appropriately used in a given context, linguistically untrained individuals are often not able to choose the form which is STRUCTURALLY equivalent to a given model. (Another manifestation of this inability to operate in structural terms is the difficulty native speakers of English taking introductory syntax have in converting an English sentence to its structurally corresponding passive and vice versa.)

3. OVERVIEW OF THE SYSTEM

To get around the difficulties outlined above a relatively simple machine-assisted translation technique was devised.

Since the transcripts of the tapes were being word processed, it was relatively easy to generate a consistent piece-by-piece (a 'piece' being a single morph or a combination of

morphs, written as an unbroken string of symbols between spaces) gloss, which could then be used by research assistants as a rough guide to their translation.

The main components of the machine glossing system are listed in figure 1. For the moment we will be concerned with only the first two of these.

In the left-most column we find a number of data files: the most important of these is the glossary, a list of morphs in Sourashtra and their equivalents in Tamil.

In order to handle the morphophonemics of Tamil verbs, the system being used also includes a database of Tamil verb stem allomorphs. The glossing program selects the appropriate allomorph according to the identity of the following suffix.

The other files in this column are lists of stems and affixes derived mechanically from the glossary and used by the morphological parser program (see below).

The heart of the system is the glossing module with two components. First a morphological parser separates stems and affixes prior to glossing; this allows for a more efficient glossary, since the inflectional forms of verbs and nouns do not all have to be listed. Subsequently, the glossing program proper takes each piece (morph or combination of morphs between spaces) of a Sourashtra text in turn, looks it up in the glossary, retrieves the Tamil equivalent and adds it to the Tamil gloss.

DATA FILES	GLOSSING	ANALYSIS	HOUSEKEEPING
*Glossary	**Morphological parser	**Equivalence-pair extractor	[File maintenance programs]
*Tamil verb database	**Glossing program		
Saurashtra stem files		**Comparison program	
Saurashtra affix files			
INPUT: (none)	text file	parallel texts	(various)
OUTPUT: (none)	text with gloss	-new items for glossary -queried items	(various)

Figure 1. System modules
Extracts of *'d files are included in Appendix 1.
Listings of **'d programs are included in Appendix 2.

4. THE ROLE OF MACHINE-GLOSSING IN THE SOURASHTRA PROJECT

The method being used to compare the structures of the two languages is as follows:

1. Each Sourashtra text is machine-glossed as outlined above.

2. The machine output is given to bilingual assistants as a basis for their translation.

3. Any instances in which the assistants feel it necessary to change the computer gloss are then examined. These may be due to:

   - an error in one of the computer programs or data files
   - an erroneous or inconsistent translation on the part of the bilingual assistants.

- a typographical or transcriptional error in the original text
- a real difference in the corresponding structures.

It is differences of the last type which are of most interest, since these indicate points of incongruity between the grammars of the two languages.

In the following sections these steps are illustrated and described in more detail.

5. A TEXT

In (1) is a short piece of a conversation which was analyzed fairly early in the course of the project. It is part of a conversation in which F1 interviews her neighbour M2 about his work. Italics indicate loanwords[1]; # indicates a break in continuity; = marks an enclitic. English glosses are added here for expository purposes.

(1) Excerpt of Sourashtra conversation (interlinear gloss added)

						Free translation
1 M2>	*time*	soLabhaar	*eeLu*	*maNi*		My hours? I go at seven in the
	time	morning	seven	o'clock		morning and after that I come
2	jii	thellaadugu:.	*tiffin*			
	having gone	after that	tiffin			back at nine o'clock to eat tiffin.
3	khattagu:	*ombadu*	*maNi*	avu		
	to eat	nine	o'clock	I will come		
4	tellaa	*pattu*	*maNi*	phiir	jii	
	Then	ten	o'clock	againhaving gone		Then at ten o'clock I go
5	dhuhaatto	*reNDu*	=di	*muuNu*		again at around two or three
	afternoon	two	maybe	three		in the afternoon or else
6	=di: nhiiyat	*oru*	*maNi*	=di:		
	maybe or else	one	o'clock	maybe		maybe one o'clock I come back
7	avu.	lhekko		rhaanaa		
	will come	account		will not be		There are no fixed hours.
8	telle.	tehellaadugu:	*reNDu*	*maNi*		
	that.	after that	two	hour		After that, I take two hours'
9	*neeram*	*rest*	khaLLi		te	
	time	rest	having taken		that	rest and after that it will
10	hellaaduk	*pattu*	*maNi*	hooy		
	after	ten	o'clock	will be		be ten o'clock before I come back.
11	avattok.					
	to come					

12 F1>	ee	mundaDi	tumi	kaay	kellee		Before this what were you doing?
	this	before	you	what	doing		
13	hodiriyo:?						
	were?						
14M2>	ee	mundaDi:	sri	krishnaa			
	this	before	Sri	Krishna			
	Ironsu	-m	hodiriyo.				Before this I was at Sri Krishna Irons.
	Irons	in	was				
15 F1>	telle	kaam	tungo	oppayoni	=yaa		
	that	work	you	did not please	Q		Didn't you like that work?
16M2>	telle	oppayosi.	moduraLi				
	that	pleased-EMPH.	owner				I certainly liked the work, but
17	beDaanu: oo	avDi	hodeyo	kaam			
	sons	come	had	work			the owner's sons had begun to work
18	*staffu*	=nu	=gu:	# *staff*	veen		there and for the staff the work#
	staff	PL	for	staff	too many		They said there were too many staff
19	hores	meNTi	oNTa	di,	thiin	tengo	
	became	QUOT	one	day,	three	people	and one day they laid three
20	hibbi	rhavaTriyo.	temaam	mi:			
	stand	had caused	in that	I			people off. I was one of them.
21	oNTeno.						
	one person						

6. GLOSSING

The results of processing this text by the glossing module can be seen in (2). The text is arranged in two columns. In the left hand column is the text of (1) after it was run through the parsing program to separate stems from their affixes. When this text was processed, the parser looked only for verbs; it has since been modified to look for nouns as well. In order to locate a verb form the program compares each word to a list of verb stems which is stored in a data file. If the first part of the word matches a verb stem, it looks at the second part and compares it to a list of suffixes stored in another data file. If this matches too, the program assumes the form is a verb, and breaks it into two parts. Thus in line 1 the last word in the left hand column, *jii* 'having gone', has been broken into the stem *ji* plus the suffix *-i* which marks the past participle. The success of the parser is limited by the lists of stems and affixes. In line 2 the form *khattagu* 'in order to eat' is not broken into stem *kha* and suffix *-ttagu* because the suffix was not at that point included in the suffix list. The list has since been modified to include this suffix. In line 11 *kellee* 'doing' has also not been broken up for similar reasons.

(2)	Transcript, following parsing	Computer-generated Tamil gloss
1 M2>	*time* soLabhaar *eeLu maNi* ji -i	*time* kaalay/kaalayla *eeLu* maNi
2	thellaadugu:. *tiffin* khattagu:	pooy -i tiffin
3	*ombadu maNi* av -u. tellaa *pattu*	*ombadu* maNi varuv -een pattu
4	*maNi* phiir ji -i dhuhaatto *reNDu*	maNi pinna pooy -i reND
5	=di: *muuNu* =di: nhiiyat *oru*	=aac=um muuN =aac=um illaTTi *oru*
6	*maNi* =di: av -u lhekko	maNi =aac=um varuv -een
7	rhaanaa telle: tehellaadugu:	irukkaadu anda/adu
8	*reNDu maNi neeram.* rest khaL -Li	reNDu maNi *neeram* rest vaang
9	te hellaaduk *pattu maNi* hoo -y	-ikk-iT-Tu ada pattu maNi
10	avattok.	irund -u
11 F1>	ee mundaDi tumi kaay kellee hod	inda niinga enna
12	-iriyo:?	irund -irukk-
13 M2>	ee mundaDi: sri krishnaa *Ironsu*	inda *Ironsu* -la
14	-m hod -iriyo.	irund -irukk-
15 F1>	telle kaam tungo oppayoni =yaa?	anda/adu veela ongaLukku =aa
16 M2>	telle oppayosi moduraLi beDaa	anda/adu modalaaLi magan-
17	-nu: oo [pause] av -Di hod -eyo	-ga oo vand -uT-Tu- irund --
18	kaam *staffu* -nu =gu: # *staff* veen	veela *staffu* -ga # staff romba
19	hores meNTi oNTa di, thiin	oru kuDukk muuNu
20	tengo hibb -i rhavaTriyo temaam	avangaLukku niND -u adile
21	mi: oNTeno	naan

In the right hand column of (2) is the output of the glossing program. If a Sourashtra piece is not located in the glossary the glossing program first looks to see if it is a loanword (marked with italic control characters) and if it is, simply takes it over as-is into the Tamil gloss. If it is not a loanword the program leaves a space in the Tamil gloss to be filled in manually.

Thus in the processing of line 1 the first word *time*, a loan from English, is taken over as is, so are the Tamil words *eeLu* and *maNi* later in the line. The verb stem *ji* 'go' is found in the glossary and replaced by *pooy*, similarly the past participle suffix *-i* is replaced by the corresponding Tamil suffix, which in this case happens to be phonologically the same. In the second line the form *khattagu* is not found in the glossary, because, as we noted earlier, the parser has failed to analyze it. A space is thus left in the Tamil gloss. Typographical or transcriptional errors such as *=gu:* (for *-gu:*) 'DATIVE' in line 18 and *meNTi* (for *menti*) 'QUOTATIVE' in line 19 were obviously not found in the glossary and were treated as unknown. Thus a space was left in the gloss.

Given the simplicity of the glossing procedure, problems with homophones[2] arise. An illustration is found in line 19 where the form *di* has been found in the glossary and the Tamil equivalent *kuDukk*, the present stem of the verb 'give' has been inserted. The problem is that the Sourashtra form *di* could be one of three homophonous items: the stem of the verb 'give', the numeral 'two' or the noun 'day'; here the appropriate gloss is 'day', not 'give'. Despite these problems and despite the fact that at the time when this text was processed, the glossary contained less than two thousand entries, a large percentage of the text was glossed. At that time, however, the results were uneven, with some passages having many more holes than in the passage chosen for illustration. At the time of writing, the glossary contains nearly 6,000 entries (approximately 60% of which are affixes and affix-combinations) and results are much more consistent.

7. TRANSLATION

The text with the computer-generated gloss is given to a bilingual assistant, who fills in the gaps and makes changes wherever necessary. One thing which the glossing program does not do is handle subject concord affixes on verbs, because the concord systems of the two languages are different. Consequently, the translators nearly always have to fill these in. While this is doubtless a nuisance, it, along with the problems created by homophones, does serve to keep the translators' attention focussed on the text and to continually remind them of the imperfections in the computer output, so that they do not become reticent about changing it.

The most important part of the analysis then follows – this is the incorporation of the translator's changes and insertions and concurrent examination of these. The results of this process can be seen in (3). At this stage errors can be quickly corrected. For example, in line 27, as we saw earlier, the quotative particle had been mistranscribed and therefore not found in the glossary; as a result no corresponding Tamil form was given by the glossing program. But since it is a common form, its apparent absence from the glossary is unusual. Thus the gap in the gloss called attention to this form and the error was discovered. Where there is a difference in structure not due to error or inconsistency, the computer gloss is not changed, but the correct translation is inserted in curly brackets. For example, in line 13, the phrase *ee mundaDi* meaning 'before this' has been machine glossed as *inda munnaaDi* and corrected to *idukku munnaaDi*. Here the Tamil postposition requires the dative form of its argument, while Sourashtra uses the nominative. Similarly, in line 14 *kaay kel -lee hod -iriyo* 'What had you been doing?' has been glossed by the computer as *enna senji kiTT irund irukk[iinga]* rather than the correct *enna senji kiTT irundiinga*. Here Sourashtra uses the past perfect form of the auxiliary, while Tamil uses only the simple past. Because curly brackets are used only to mark such discrepancies, they will be easy to locate and cull out at a later stage of the analysis.

(3) **Parallel texts following manual translation**

1 M2>	*time* soLabhaar *eeLu maNi* ji -i	*time* kaalayla eeLu maNi pooy -i
2	thellaadugu:. *tiffin* kha	ad-ukk+appram *tiffin* sappiD
3	-attagu: *ombadu maNi* av	-r=ad-ukku ombadu maNi varuv
4	-u tellaa *pattu maNi* phiir	-een. anda pattu maNi pinna
5	ji -i dhuhaatto *reNDu* =di:	pooy -i mattyaaNam reND =aac=um
6	*muuNu* =di: nhiiyat oru	muuNu =aac=um illaaTTi oru
7	*maNi* =di: av -u	maNi =aac=um varuv -een.
8	lhekko rhaanaa telle:	kaNakku irukkaadu adu.
9	tehellaadugu: *reNDu maNi neeram*	ad -ukku+peragu reNDu maNi neeram
10	*rest* khaL -Li te	*rest* eDutt -ukk-iT-Tu ad
11	hellaaduk *pattu maNi* hoo -y	-ukku+peragu pattu maNi aagum
12	av -attok	var -r=ad-ukku.
13 F1>	ee mundaDi tumi	inda {idukku} munnaaDi niinga
14	kaay kel -lee hod -iriyo:?	enna senj -ikkiTT
15		irund -irukk-iinga {senj -ikkiTT
16		irund -iinga}?
17 M2>	ee mundaDi: sri	inda {idukku} munnaaDi, sri
18	kriSnaa *Ironsu* -m hod	kriSnaa *Irons* -la irund
19	-iriyo.	-irukk-een.
20 F1>	telle kaam tungo opp	anda veela ongaLukku piDikk
21	-ayoni =yaa?	-ala =aa?
22 M2>	telle oppayo =si	adu piDicc -adu =taan.
23	moduraLi beDaa -nu: oo	modalaaLi magan- -unga oo
24	[pause] av -Di hod	[pause] vand -uT-Tu- irund
25	-eyo kaam *staffu* -nu	-aanga. veela *staffu* -gaL
26	-gu: # *staff* veen hores	-ukku # *staff* adigam aagudu
27	menti, oNTa di thiin	=n-iT-T-u, oru naaL muuNu
28	=tengo hibbi+rhava	=peer-ukku nippaaTT
29	-Triyo temaam mi: oNTeno	-iTT-aanga. ad-ile naan orttan

As a final stage in the processing, each text was submitted to a linguistically trained native speaker of Tamil to verify the naturalness of the translation. At this stage errors caused by the bilingual speakers' imperfect knowledge of Tamil and/or interference from Sourashtra can be corrected. For example in line 1, *eeLu maNi pooyi*, 'having gone at 7 o'clock' would be corrected to *eeLu maNikku pooyi*, since Tamil requires the dative case in this construction.

8. PROCESSING TIME

The transcript of a half hour tape (approximately 1000 lines) takes about 45 minutes to one hour to parse and 25 to 30 minutes to gloss using an IBM-AT compatible computer running at 12 MHz. The times are slow because all the programs are run in (compiled) BASIC rather than a more efficient language. This is because I know BASIC best and can get programs debugged and up and running faster in this language. With only about 70 half-hour tapes to process, computer time used is not a prime concern, particularly given the vast quantities of human time required to process the texts[3].

9. THE ANALYSIS MODULE

Let us now return to the analysis module in figure 1. The input to this module is two parallel texts which have been fully processed and checked. We will use the term EQUIVALENCE-PAIR for a Sourashtra piece and its corresponding Tamil translation in such a text. An equivalence-pair extraction program makes a list of such pairs by taking each piece in the Sourashtra text and matching it with the piece in the same position in the Tamil text (ignoring bracketed material). The procedure relies on the fact that word-order in the two languages is virtually identical. Thus in our example, *time* would be matched with *time*, *soLabhaar* with *kaalayla*, *eeLu* with *eeLu* and so on. The resulting list is shown in part in (4). A second program sorts these pairs and compares each one with the glossary, producing two lists: one of words not found in the glossary and another of words found in the glossary but with a different gloss from the one in the text. The first of these is shown in (5).

The checking of these lists helps further reduce the number of typographical errors in the text files. For example, errors in the input text sometimes cause an extra piece (unbroken stretch of symbols between spaces) to appear in one of the two columns. Thus in line 22 the Sourashtra *oppayo* 'pleasePAST' has been translated by two separate pieces *piDicc* 'pleasePAST' and *-adu* '3sgNONHUMAN'. In such cases, the equivalence-pair extraction program blindly continues to match pieces by their corresponding position, producing in our example pairs such as *=si* 'EMPHATIC'/*-adu* '3sgNONHUMAN', *moduraLi* 'boss'/*=taan* 'EMPHATIC' etc. At the end of the speaker's turn, however, the program discovers an unequal number of pieces in the two columns. When this happens it throws away the pairs produced in the current turn and sends a message noting the discrepancy to an error file.

Occasionally a pair of errors, one in each column, cancel each other's effect, so that an equal number of pieces appears in both columns. A (quite fortuitous) instance of this type of double error appears in (3): in line 9, Sourashtra *tehellaadugu* 'afterwards' is translated as *ad -ukku+peragu* 'that DAT+after'; then in line 11 Sourashtra *hoo -y* 'be -FUTUREsg' is translated by the single piece *aagum* 'be:FUTURE:3sgNONHUMAN'. Such double errors go unnoticed by the equivalence-pair extraction program and are passed on to the glossary comparison program, where their treatment will depend upon whether the Sourashtra member of a pair is in the glossary. If it is, then its Tamil equivalent there will differ from the Tamil member of the equivalence-pair, and the program will signal this fact in its list of queried pairs. Thus in our example, the loan *reNDu* 'two' from the Sourashtra column paired with *-ukku+peragu* 'DAT+after' will clash with the glossary entry *reNDu*/reNDu. If the Sourashtra member of an erroneous equivalence-pair is not in the glossary, the pair will be included in the list of items to be added to the glossary; see for

example the pair *rest* 'rest'/*neeram* 'time' in (5). Such pairs may be culled out when the list is verified, before being merged into the glossary. If the glossary is of a reasonable size, it is likely that at least one of the erroneous equivalence-pairs in a double-error situation such as this will indeed be in the glossary, so that it will not be necessary to rely solely on checking the 'to be added' list in order to locate the error.

10. CONCLUSIONS

The use of computer glossing in this study has succeeded in ensuring that linguistically untrained bilinguals produce consistent and accurate translations, so that a large corpus of texts can be reliably processed. The technique also helps to locate points of incongruity in the structure of the two languages and incidentally helps locate transcriptional inaccuracies and typographical errors in the initial text.

The use of a machine-generated gloss to guide translation is possible because the languages involved share a great many structural features and semantic categories, and consequently the glosses are a good approximation to the final translations. The approach is thus ideally suited for work in convergence situations.

Finally the system requires only simple programming techniques which are not beyond the reach of the computer-literate linguist.

(4) **Partial output of equivalence-pair extraction program**

time	*time*	-attagu,	-r=ad-ukku
soLabhaar,	kaalayla	*ombadu*,	ombadu
eeLu,	eeLu	*maNi*,	maNi
maNi,	maNi	av,	varuv
ji,	pooy	-u,	-een
-i,	-i	tellaa,	anda
thellaadugu,	ad-ukk+appram	*pattu*,	pattu
tiffin,	*tiffin*	*maNi*,	maNi
kha,	sappiD	phiir,	pinna

(5)	**Output of glossary comparison program**		
Ironsu	,Ironsu	lhekko	,kaNakku
eeLu	,eeLu	mundaDi	,munnaaDi
neeram	,maNi	sri	,sri
ombadu	,ombadu	tehellaadugu	,ad
oru	,oru	tellaa	,anda
rest	,neeram	thellaadugu	,ad-ukk+appram
dhuhaatto	,mattyaaNam	-attagu	,-r=ad-ukku
hellaaduk	,ad	-attok	,-r=ad-ukk
kriSnaa	,kriSnaa	-ayoni	,-ala

ACKNOWLEDGEMENTS

This paper stems from research funded by the Social Sciences and Humanities Research Council of Canada, the Shastri Indo-Canadian Institute, and York University. The author also wishes to acknowledge the advice and logistical support of J. Neethivanan and the invaluable research assistance provided by A. R. Kumar, N. M. Omprakash, R. R. Parimalam, K. Pasumpon, P. Rajendran, O. K. Ramanandam, V. V. Ramlakshmanji, and K. S. Santhi. Earlier versions of this paper have been read at a Seminar on Society and Language, Madurai 1989 and the Seminar on Development of Common Core Grammar of Indian Languages for Computers, Mysore, 1990.

FOOTNOTES

1 The term LOANWORD is being used to cover both established loans, for which a Sourashtra equivalent does not exist or is rarely used as well as more ephemeral loans for which a Sourashtra equivalent is in common use.

2 Pre-coding the input text to distinguish among homophones was rejected as being too time consuming.

3 In the Sourashtra-Tamil convergence study, transcription of a half-hour conversation took between 24 and 40 hours; translation required a similar amount of time. Initial checking each of these operations required another 8 to 15 hours.

APPENDIX 1

EXTRACTS OF DATA FILES

1. GLOSSARY

Sourashtra	Tamil	Sourashtra	Tamil
virtaa	, viiNaa	-DNaar	, -ikk-r=avanga
virudunagar	, virudunagarN	-DNo+poDariyo:	, -iD-a+veeNDiy-adu
visayam	, visayamN	-DaDatte	, -iD-a+vekk-r=adu
visees	, viseesamN	-Dadi	, -iT-T-aa
viseesu	, viseesamN	-Dadi+meLLi	, -iT-T-aal=um
visi+laga	, salipp=aa	-Dadi+meLi	, -iT-T-aal=um
visir	, viisV	-Dan	, -iD-uv-
visnu	, visnuN	-Dan+jaariyo	, -iD-a+poo-r-
vistaay	, viisi+pooDu	-Dan+jaas	, -iD-a+ppoogudu
vistag	, viisi+pooDV	Danda	, -iD-aTTum
vistav	, viisi+pooDV	-Dando	, -iD-aTTum
vistavo	, viisi+pooDunga	-Danu	, -iD-uv-
vistay	, viisi+pooDV	-Daraas	, -iD-r-

2. TAMIL VERB DATABASE

(pres stem)	(inf. stem)	(past stem)	(past part)	(fut stem)
irukk	,irukk	,irund	,irundu	,irupp
piDikk	,piDikk	,piDicc	,piDicci	,piDipp
poo	,poog	,poon	,pooyi	,poov
saappiD	,saappiD	,saappiTT	,saappiTTu	,saappiDuv
vaang	,vaang	,vaangin	,vaangi	,vaanguv
var	,var	,vand	,vandu	,varuv

APPENDIX 2

PROGRAM LISTINGS

The program listings included herein are current working versions, which have been edited only to break up overlength lines. The programs could be more modular in design, but since any modifications would require testing and debugging, for which there is not time, it is preferable to give programs which do actually work.

1. PARSING PROGRAM

```
100 '   PARSER.BAS
110 '
120 '   THIS VERSION: February 22, 1990
121 '
122 '   IAN SMITH, LINGUISTICS, YORK
        UNIVERSITY, NORTH YORK, ONT,
        CANADA M3J 1P3
123 '
130 '   PARSES SOURASHTRA VERB FORMS
        INTO ROOT PLUS A SINGLE SUFFIX-
        COMBINATION
135 '   N.B. THIS VERSION REQUIRES A DRIVE
        D (RAMDISK) WHICH IS NOT THE
        DEFAULT
136 '
140 '
500 '   *** FILES ***
510 '
520 '   VAFFIXES.SAU  ASCII-SORTED LIST OF
        AFFIX-COMBINATIONS OF
        SAURASHTRA
530 '   VERBS. RANDOM. RECORD 20.
540 '   filename  TEXT FILE TO BE INPUT.
        SAURASHTRA TEXT IN COLUMNS 1-41.
545 '   LINE BEGINNING CHR$(11)+CH$(15)
        SEPARATES TURNS.
550 '   PARSER.OUT  OUTPUT FILE,
        FORMATTED AS ABOVE WITH VERB
        FORMS
570 '   PARSER.ERR  ERROR FILE
580 '   VSTEMS.SAU  ASCII-SORTED LIST OF
        SAURASHTRA VERB STEMS. RANDOM
585 '     RECORD 20.
590 '   VSTEMS.DAT  DATA FOR
        VSTEMS.SAU: NUMBER OF ITEMS
600 '   VAFFIXES.DAT  DATA FOR
        VAFFIXES.SAU: NUMBER OF ITEMS
610 '   NSTEMS.SAU  ASCII-SORTED LIST OF
        SAURASHTRA NOUN STEMS. RANDOM
620 '     RECORD 20.
630 '   NSTEMS.DAT  DATA FOR
        NSTEMS.SAU: NUMBER OF ITEMS
640 '   NAFFIXES.DAT  DATA FOR
        NAFFIXES.SAU: NUMBER OF ITEMS
690 '
700 '   *** VARIABLES ***
710 '
730 '   W$  SAURASHTRA WORD OF A
        PARTICULAR TURN - NET, AFTER
740 '     PARENTHETICAL MATERIAL ETC
        REMOVED.
742 '   STEM$  VERB STEM IDENTIFIED IN W$
744 '   RMP$  REMAINDER OF W$
        (=POTENTIAL AFFIX COMPLEX)
750 '   LO$( )  OUTPUT LINES OF A TURN
770 '   LI  LINE NUMBER COUNTER
790 '   ILK  COUNTER FOR INPUT LINES IN
        TURN
800 '   OLK  COUNTER FOR OUTPUT LINES IN
        TURN
810 '   ASIZ  NUMBER OF ITEMS IN
        "VAFFIXES.SAU"
815 '   IC  COUNTER FOR ITALICS CONTROL
        CHARACTERS
820 '   VSIZ  NUMBER OF ITEMS IN
        "VSTEMS.SAU"
825 '   NASIZ  NUMBER OF ITEMS IN
        "NAFFIXES.SAU"
830 '   NSIZ  NUMBER OF ITEMS IN
        "NSTEMS.SAU"
850 '   PTR  POINTER TO POSITION IN
        CURRENT LINE
860 '   BREND  POINTER TO POSITION AT
        WHICH BRACKET CLOSES
870 '   BR$  BRACKET OPEN FLAG/TYPE OF
        RIGHT BRACKET EXPECTED
880 '   BEG  POINTER TO POSITION IN LINE AT
        WHICH WORD BEGINS
890 '   F$  LEFTMOST CHARACTER OF
        SOMETHING
900 '   MCH  FLAG FOR GOOD MATCH (I.E.
        STEM FROM FILE = STEM$)
910 '   LST  LENGTH OF STEM$
920 '   B.VSTEM$  BUFFER FOR VSTEMS.SAU
        ON RAMDISK
925 '   B.NSTEM$  BUFFER FOR NSTEMS.SAU
        ON RAMDISK
930 '   B.W$  BUFFER FOR
        VSTEMS.SAU/VAFFIXES.SAU ON
        DEFAULT DRIVE
940 '   B.VAFF$  BUFFER FOR VAFFIXES.SAU
        ON RAMDISK
945 '   B.NAFF$  BUFFER FOR NAFFIXES.SAU
        ON RAMDISK
950 '   MDL  POINTER TO LOCATION IN
        BINARY SEARCH (AND AFTERMATH)
960 '
```

```
1000 ' *** PROGRAMME ***
1060    DIM LO$(500)
1100    REM INITIALIZE COUNTERS AND FLAGS
1110     LI=0:LTMP=0
1112    INPUT "NAME OF FILE TO PROCESS:
        ",F$
1114     OPEN "I",#1,F$
1116    OPEN "R",#2,"D:VSTEMS.SAU",20
1118    FIELD #2, 20 AS B.VSTEM$
1120    OPEN "I",#3,"VSTEMS.DAT"
1122    INPUT#3, VSIZ     'GET SIZE OF VERB
        STEMS FILE
1124    CLOSE #3
1126    OPEN "R",3,"VSTEMS.SAU",20
1128    FIELD #3, 20 AS B.W$
1130     PRINT "TRANSFERRING VSTEMS TO
        RAMDISK ..."
1132       FOR I=1 TO VSIZ
1134     GET#3
1136     LSET B.VSTEM$=B.W$
1140     PUT#2
1145    NEXT I
1155    CLOSE #3
1160    PRINT "VSTEMS SUCCESSFULLY
        TRANSFERRED.  NO. OF ENTRIES =
        ";VSIZ
1161    OPEN "I",#3, "VAFFIXES.DAT"
1162    INPUT#3, ASIZ
1163     CLOSE #3
1164    OPEN "R",3,"D:VAFFIXES.SAU",20
1165    FIELD #3, 20 AS B.VAFF$
1166    OPEN "R",#4,"VAFFIXES.SAU",20
                'OPEN AFFIX FILE
1168    FIELD #4, 20 AS B.W$
1170    PRINT "TRANSFERRING VAFFIXES TO
        RAMDISK ..."
1172    FOR I=1 TO ASIZ
1174     GET#4
1176     LSET B.VAFF$=B.W$
1180     PUT#3
1185    NEXT I
1195    CLOSE #4
1200     PRINT "VAFFIXES SUCCESSFULLY
        TRANSFERRED.  NO. OF ENTRIES =
        ";ASIZ
1216    OPEN "R",#6,"D:NSTEMS.SAU",20
1218    FIELD #6, 20 AS B.NSTEM$
1220    OPEN "I",#7,"NSTEMS.DAT"
1222    INPUT#7, NSIZ     'GET SIZE OF VERB
        STEMS FILE
1224     CLOSE #7
1226    OPEN "R",#7,"NSTEMS.SAU",20
1228    FIELD #7, 20 AS B.W$
1230    PRINT "TRANSFERRING NSTEMS TO
        RAMDISK ..."
1232    FOR I=1 TO NSIZ
1234     GET#7
1236     LSET B.NSTEM$=B.W$
1240     PUT#6
1245    NEXT I
1255    CLOSE #7
1260     PRINT "NSTEMS SUCCESSFULLY
        TRANSFERRED.  NO. OF ENTRIES =
        ";NSIZ
1261     OPEN "I",#7, "NAFFIXES.DAT"
1262     INPUT#7, NASIZ
1263     CLOSE #7
1264    OPEN "R",#7,"D:NAFFIXES.SAU",20
1265     FIELD #7, 20 AS B.NAFF$
1266     OPEN "R",#4,"NAFFIXES.SAU",2
        'OPEN AFFIX FILE
1268    FIELD #4, 20 AS B.W$
1270    PRINT "TRANSFERRING NAFFIXES TO
        RAMDISK ..."
1272    FOR I=1 TO NASIZ
1274     GET#4
1276     LSET B.NAFF$=B.W$
1280     PUT#7
1285    NEXT I
1295    CLOSE #4
1300     PRINT "NAFFIXES SUCCESSFULLY
        TRANSFERRED.  NO. OF ENTRIES =
        ";NASIZ
1320    INPUT "NAME OF OUTPUT FILE
        (DEFAULT IS PARSER.OUT): ",F$
1321    IF F$="" THEN F$="PARSER.OUT"
1330    OPEN "O",#4,F$
1335    OPEN "O",#5,"PARSER.ERR"
1349    REM START NEW TURN
1350    REM RESET COUNTERS
1351    I=0
1352    J=0
1353    ILK=0
1355     BR$=""   'INITIALIZE FLAG
1400     IF EOF(1) GOTO 9000
1401    IF LTMP=50 THEN PRINT LI;" LINES
        PROCESSED":LTMP=0
1402     LINE INPUT#1,L$
1403    IC=0:PTR=0   'SET COUNTERS
1405    LN=LEN(L$):RM=LN
1410     REM CHECK FOR COL BRK
1415    IF LEFT$(L$,2)=CHR$(11)+CHR$(15)
        GOTO 8280
1417    IF LEFT$(L$,1)=CHR$(12) GOTO
1400    'IGNORE PAGE BREAK 1418  'SKIP CTRL
        LINES (BUT THEY WILL BE SHIFTED TO
        BEGINNING OF TURN IN O/P FILE)
1419    IF LEFT$(L$,1)=CHR$(11) THEN
        PRINT#4,L$:GOTO 1351
1420
        LI=LI+1:LTMP=LTMP+1:ILK=ILK+1:LO$
        (ILK)=L$
1460     IF BR$<>"" THEN GOSUB 7730:GOTO
        1550
1540     'RESET LENGTH AND CHECK FOR EOL
1550    RM=LEN(L$)
1555    IF RM<1 GOTO 1400   'IF 1, THEN MUST
        BE SPACE
1559     'CHECK FOR BRACKET CONDITION OR
        BEGINNING OF BRACKET
1562    F$=LEFT$(L$,1)
1563    IF F$="<" OR F$="{" OR F$="[" THEN
        GOSUB 7600:GOSUB 7730:GOTO 1550
1570     REM IGNORE LEADING SPACES AND
        "(" AND "|"
1575    IF F$=" " OR F$="(" OR F$="|" THEN
        L$=RIGHT$(L$,RM-
        1):PTR=PTR+1:GOTO 1550
```

```
1600    REM CAPTURE WORD
1610    GOSUB 7000
1700    REM CHECK FOR NULL WORD
1710    IF W$="" GOTO 1550
1719    ' *** VERB SECTION ***
1720    REM CHECK FOR MATCH IN ROOT
1730    GOSUB 6200
1740    IF MCH=0 GOTO 1830
1770 '  STEM MATCHES. CHECK FOR AFFIX
        MATCH
1780    RMP$=RIGHT$(W$,LEN(W$)-LST)
1790    GOSUB 6000
1800    IF
        RMP$<>LEFT$(B.VAFF$,INSTR(B.VAFF
        $," ")-1) GOTO 1820
1810    LO$(ILK)=LEFT$(LO$(ILK),BEG+LST)+"
        "+CHR$(246) +RIGHT$ (LO$(ILK), LN-
        BEG-LST)
1815    LN=LN+2:PTR=PTR+2:GOTO 1550
1819 '  *** NOUN SECTION ***
1820    REM CHECK FOR MATCH IN ROOT
1830    GOSUB 5200
1840    IF MCH=0 GOTO 1550
1870 '  STEM MATCHES. CHECK FOR AFFIX
        MATCH
1880    RMP$=RIGHT$(W$,LEN(W$)-LST)
1890    GOSUB 5000
1900    IF
        RMP$<>LEFT$(B.NAFF$,INSTR(B.NAFF
        $," ")-1) GOTO 1550
1910    LO$(ILK)=LEFT$(LO$(ILK),BEG+LST)+"
        "+CHR$(126) +RIGHT$(LO$(ILK),LN-
        BEG-LST)
1920    LN=LN+2:PTR=PTR+2
1930    GOTO 1550
2030 '
4999 '  *** SUBROUTINE FOR BINARY SEARCH
        OF NAFFIXES *****
5000    UL=NASIZ-1   'last entry is dummy, so not
        NASIZ
5010    LL=1
5020    MDL=INT((UL+LL+1)/2)
5030    GET#7,MDL
5040    WHILE RMP$<>B.NAFF$ AND MDL<>UL
5050    IF RMP$<B.NAFF$ THEN UL=MDL ELSE
        LL=MDL
5080    MDL=INT((UL+LL+1)/2)
5090    GET#7,MDL
5100    WEND
5120    RETURN
5199 '  *** SUBROUTINE FOR BINARY SEARCH
        OF NOUN FILE ***
5200    UL=NSIZ   'normally size+1, but here last
        elt is dummy
5210    LL=1
5220    MDL=INT((UL+LL)/2)
5230    GET#6,MDL
5240    WHILE W$<>B.NSTEM$ AND MDL<>LL
5250    IF W$<B.NSTEM$ THEN UL=MDL ELSE
        LL=MDL
5260    MDL=INT((UL+LL)/2)
5270    GET#6,MDL
5280    WEND
5289 '  FIND OUT IF B.NSTEM$ IS A SUBTRING
        OF W$
5290
                STEM$=LEFT$(B.NSTEM$,INST
        R(B.NSTEM$," ")-1)
5300    LST=LEN(STEM$)
5310    MCH=0
5320    WHILE LST>1 AND MCH=0   'ASSUMES
        NO STEMS OF LENGTH 1
5330    IF STEM$=LEFT$(W$,LST) THEN
        MCH=1:GOTO 5500
5340    LST=LST-1
5350 '  NO MATCH: CRAWL BACK TO NEXT
        SHORTEST STEM
5360    WHILE MDL>1 AND LEN(STEM$)>LST
5370    MDL=MDL-1
5380    GET#6,MDL
5390
        STEM$=LEFT$(B.NSTEM$,INSTR(B.NST
        EM$," ")-1)
5400    WEND
5500    WEND
5510    RETURN
5998 '
5999 '  *** SUBROUTINE FOR BINARY SEARCH
        OF VAFFIXES ***
6000    UL=ASIZ-1   'last entry is dummy, so not
        ASIZ
6010    LL=1
6020    MDL=INT((UL+LL+1)/2)
6030    GET#3,MDL
6040    WHILE RMP$<>B.VAFF$ AND MDL<>UL
6050    IF RMP$<B.VAFF$ THEN UL=MDL ELSE
        LL=MDL
6080    MDL=INT((UL+LL+1)/2)
6090    GET#3,MDL
6100    WEND
6120    RETURN
6199 '  *** SUBROUTINE FOR BINARY SEARCH
        OF VERB FILE ***
6200    UL=VSIZ   'normally size+1, but here last
        elt is dummy
6210    LL=1
6220    MDL=INT((UL+LL)/2)
6230    GET#2,MDL
6240    WHILE W$<>B.VSTEM$ AND MDL<>LL
6250    IF W$<B.VSTEM$ THEN UL=MDL ELSE
        LL=MDL
6260    MDL=INT((UL+LL)/2)
6270    GET#2,MDL
6280    WEND
6289 '  FIND OUT IF B.VSTEM$ IS A SUBTRING
        OF W$
6290
                STEM$=LEFT$(B.VSTEM$,INST
        R(B.VSTEM$," ")-1)
6300    LST=LEN(STEM$)
6310    MCH=0
6320    WHILE LST>1 AND MCH=0   'ASSUMES
        NO STEMS OF LENGTH 1
6330    IF STEM$=LEFT$(W$,LST) THEN
        MCH=1:GOTO 6500
6340    LST=LST-1
```

```
6350 '   NO MATCH: CRAWL BACK TO NEXT
         SHORTEST STEM
6360     WHILE MDL>1 AND LEN(STEM$)>LST
6370      MDL=MDL-1
6380      GET#2,MDL
6390
         STEM$=LEFT$(B.VSTEM$,INSTR(B.VST
         EM$," ")-1)
6400      WEND
6500     WEND
6510     RETURN
6998 '
6999 '   ******* WORD CAPTURE SUBROUTINE
         *********
7000     W$="":BEG=PTR
7005     WHILE LEFT$(L$,1)<>" " AND RM>0
7006     REM IGNORE PUNCTUATION
7007      F$=LEFT$(L$,1):PTR=PTR+1
7012      IF F$="." OR F$="," OR F$="?" OR
         F$=":" OR F$=";" OR F$="!" GOTO 7020
7013      IF F$="!" OR F$="(" OR F$=")" OR
         F$="#" GOTO 7020
7014     IF F$="]" OR F$="}" OR F$=">" THEN
         PRINT#5, "UNPAIRED ";F$;" IN LINE
         ";LI:GOTO 7020
7015     W$=W$+F$
7016     IF F$=CHR$(21) THEN IC=IC+
                  'INCREMENT ITALIC CTRL
         CHAR COUNTER
7020     RM=RM-1
7025     L$=RIGHT$(L$,RM) 7030  WEND
7050      RETURN
7060 '
7599 '   *** SUBROUTINE TO SET BRACKET
         FLAG ***
7600     IF F$="[" THEN BR$="]"
7650     IF F$="{" THEN BR$="}"

7685     IF F$="<" THEN BR$=">"
7700     RETURN
7701 '
7728 '
7729 '   *** SUBROUTINE TO IGNORE
         BRACKETTED MATERIAL ***
7730     BREND=INSTR(L$,BR$)
7740     IF BREND=0 THEN L$="": GOTO 7780
7750     L$=RIGHT$(L$,RM-BREND)
7760     BR$=""
7770     PTR=PTR+BREND
7780      RETURN
7998 '
8199 '   *** OUTPUT A TURN ***
8280     FOR K=1 TO ILK   'LO$(ILK) = line before
         COLBRK
8290     PRINT#4,LO$(K)
8300     NEXT K
8310     PRINT#4,CHR$(11);CHR$(15)   'PRINT
         COLUMN BREAK
8320     IF BR$<>"" THEN PRINT#5, "MISSING
         ";BR$;" IN LINE ";LI
8325     GOTO 1351
8330 '
9000      CLOSE
9010      PRINT "OPERATION SUCCESSFULLY
         TERMINATED"
9011      BEEP:BEEP:BEEP:BEEP:BEEP
9020      END
```

2. GLOSSING PROGRAM

```
100 '    GLOSS.BAS
110 '
120 '    THIS VERSION: March 26, 1990
122 '    IAN SMITH, LINGUISTICS, YORK
         UNIVERSITY, NORTH YORK, ONT.
         CANADA M3J 1P3
124 '
130 '    GLOSSES FROM SOURASHTRA TO
         TAMIL ON A WORD-BY-WORD BASIS
135 '    15-11-89 UPDATE ADDS CAPACITY FOR
         HANDLING VERB MORPHOLOGY AND
         SOME
136 '    SANDHI PROCESSES. 22-2-90 VERSION
         ADDS CAPACITY FOR PARSING NOUNS
137 '    N.B. THIS VERSION REQUIRES
         RAMDISK D: (NOT DEFAULT) WHICH IS
         LARGE
138 '    ENOUGH TO HOLD GLOSSARY AND
         ANCILLARY DATA FILES.
140 '
500 '    ************ FILES ************
510 '
520 '    GLOSSARY.ST  ASCII-SORTED
         SOURASHTRA-TAMIL GLOSSARY.
         RANDOM FILE
530 '    RECORD 40: SOURASTRA WORD=20;
         TAMIL WORD=20
540 '    filename   TEXT FILE TO BE INPUT.
         SOURASHTRA TEXT IN COLUMNS 1-41.
545 '    LINE BEGINNING CHR$(11)+CH$(15)
         SEPARATES TURNS.
550 '    GLOSS.TXT        OUTPUT FILE,
         FORMATTED AS ABOVE WITH TAMIL
         ADDED IN
560 '    COLUMNS 47-77
570 '    GLOSS.ERR   ERROR FILE
580 '    GLOSSARY.DAT   DATA FOR
         GLOSSARY.ST: NUMBER OF ITEMS
590 '    VBASE.TAM   ASCII-SORTED
         PRINCIPAL PARTS OF TAMIL VERBS.
         RANDOM.
600 '    RECORD 75 = 5X15:
610 '    PRES STEM, INF STEM, PAST STEM,
         PAST PPL, FUT STEM
690 '
700 '    *********** VARIABLES ***********
710 '
715 '    F$        I/O filename
720 '    SW$( )  SOURASHTRA WORDS OF A
         PARTCULAR TURN - GROSS
730 '    SO$( )  SOURASHTRA WORDS OF A
         PARTICULAR TURN - NET, AFTER
```

```
740 '    PARENTHETICAL MATERIAL ETC
         REMOVED.
750 '    LO$( )  OUTPUT LINES OF A TURN
760 '    TU$( )  TAMIL WORDS
         CORRESPONDING TO SO$( ).
770 '    LI  LINE NUMBER COUNTER
790 '    ILK  COUNTER FOR INPUT LINES IN
         TURN
800 '    OLK  COUNTER FOR OUTPUT LINES IN
         TURN
810 '    GSIZ  NUMBER OF ITEMS IN
         "GLOSSARY.ST"
820 '    IC  COUNTER FOR ITALICS CONTROL
         CHARACTERS
830 '    TW$  TAMIL WORD WITH PADDING,
         INPUT TO MORPHOLOGICAL SEARCH
840 '    VSIZ  NUMBER OF ITEMS IN
         "VBASE.TAM"
850 '    STM$  APPROPRIATE VERB STEM,
         PADDED WITH SPACES
860 '
1000 '   *** PROGRAMME ***
1055     DIM SW$(2500)
1057     DIM SO$(2500)
1060     DIM LO$(500)
1070     DIM TU$(2500)
1100      REM INITIALIZE COUNTERS AND
         FLAGS
1110     LI=0:LTMP=0
1112      INPUT "NAME OF FILE TO PROCESS:
         ",F$
1114      OPEN "I",1,F$
1116     OPEN "R",2,"D:GLOSSARY.ST",40
1118     FIELD #2, 20 AS B.SGW$,20 AS B.TGW$
1120     OPEN "I",3,"GLOSSARY.DAT"
1122      REM INPUT GLOSSARY SIZE
1124     INPUT#3, GSIZ
1125     CLOSE #3
1126     OPEN "R",3,"GLOSSARY.ST",40
1128     FIELD #3, 20 AS B.SW$,20 AS B.TW$
1130      PRINT "TRANSFERRING GLOSSARY TO
         RAMDISK ..."
1132     FOR I=1 TO GSIZ
1134      GET#3
1136      LSET B.SGW$=B.SW$
1138      LSET B.TGW$=B.TW$
1140      PUT#2
1145     NEXT I
1155     CLOSE #3
1160      PRINT "GLOSSARY SUCCESSFULLY
         TRANSFERRED.  NO. OF ENTRIES =
         ";GSIZ
1200      OPEN "I",3,"VBASE.DAT"
1202      INPUT#3, VSIZ   'GET SIZE OF VERB
         FILE
1204      CLOSE #3
1206      OPEN "R",3,"VBASE.TAM",75
                  'OPEN VERB FILE
1208      FIELD #3, 15 AS B.PRS$,15 AS
         B.INF$,15 AS B.PST$,15 AS B.PPL$,15
         AS B.FUT$
1220     INPUT "NAME OF OUTPUT FILE
         (DEFAULT IS GLOSS.TXT): ",F$
1221     IF F$="" THEN F$="GLOSS.TXT"
1230     OPEN "O",4,F$
1235     OPEN "O",5,"GLOSS.ERR"
1299      REM START NEW TURN
1300      REM RESET COUNTERS
1351     I=0
1352     J=0
1353     ILK=0 1400  IF LTMP=50 THEN PRINT
         LI;" LINES PROCESSED":LTMP=0
1401      IF EOF(1) GOTO 9000
1402      LINE INPUT#1,L$
1403     IC=0   'SET COUNTER
1405     LN=LEN(L$)
1410      REM CHECK FOR COL BRK
1415     IF LEFT$(L$,2)=CHR$(11)+CHR$(15)
         GOTO 7565
1417     IF LEFT$(L$,1)=CHR$(12) GOTO 1400
         'IGNORE PAGE BREAK
1418     'SKIP CTRL LINES (BUT THEY WILL BE
         SHIFTED TO BEGINNING OF TURN IN O/P
         FILE
1419     IF LEFT$(L$,1)=CHR$(11) THEN
         PRINT#4,L$:GOTO 1351
1420
         LI=LI+1:LTMP=LTMP+1:ILK=ILK+1:LO$
         (ILK)=L$
1435      IF LN<8 GOTO 1400
1540      REM RESET LENGTH AND CHECK FOR
         EOL
1550      RM=LEN(L$)
1560      IF RM<1 THEN GOSUB 2000:GOTO 1400
         'IF 1, THEN MUST BE SPACE
1570      REM REMOVE LEADING SPACES
1575     IF LEFT$(L$,1)=" " THEN
         L$=RIGHT$(L$,RM-1): GOTO 1550
1600      REM CAPTURE WORD
1610     GOSUB 7000
1700      REM CHECK FOR NULL WORD
1710     IF W$="" GOTO 1540
1720      REM  ASSIGN TO SW ARRAY
1730     I=I+1:SW$(I)=W$:GOTO 1550
1998 '
1999 '   *** SUBROUTINE TO PAD/TRUNCATE
         LINE TO DIVIDING POINT ***
2000      IF LEN(LO$(ILK))>45+IC THEN
         PRINT#5,"MATERIAL BEYOND COLUMN
         ";45+IC; " IN LINE ";LI;"
         TRUNCATED":LO$(ILK)
         =LEFT$(LO$(ILK),45+IC):LO$(ILK)=LEF
         T$(LO$(ILK),45+IC)
2010      LO$(ILK)=LO$(ILK)+SPACE$(46+IC-
         LEN(LO$(ILK)))
2020      RETURN
2030 '
4999 '   *** SANDHI PROCESSOR ***
5000     FOR K=1 TO IO-1
5010      IF RIGHT$(TU$(K),2)<>"u " OR
         RIGHT$(TU$(K),3)="uu " GOTO 5050
5020     ONS$=LEFT$(TU$(K+1),2)
5030     IF ONS$="-i" OR ONS$="-a" OR ONS$="-
         o" OR ONS$="-e" OR ONS$="=i" OR
         ONS$="=a" OR ONS$="=o" OR
         ONS$="=e"THEN
         TU$(K)=LEFT$(TU$(K),LEN(TU$(K))-2)+"
         "
```

```
5050    IF RIGHT$(TU$(K),2)<>"a " GOTO 5090
5060    ONS$=LEFT$(TU$(K+1),2)
5070    IF ONS$="-i" OR ONS$="-a" OR
        ONS$="-o" OR ONS$="-e" OR ONS$="-u"
        OR ONS$="=i" OR ONS$="=a" OR
        ONS$="=o" OR ONS$="=e" OR
        ONS$="=u"  THEN
        TU$(K)=LEFT$(TU$(K),LEN(TU$(K))-
        1)+"v "
5090    NEXT K
5999 '  *** SUBROUTINE: MORPHOLOGICAL
        SUBCOMPONENT ***
6000    K=1
6001    WHILE K<=IO
6019 '  ========= VERB SECTION =========
6020    WHILE
        RIGHT$(TU$(K),3)="V"+CHR$(2)+" "
6025    TU$(K)=LEFT$(TU$(K),LEN(TU$(K))-4)
        'CUT OFF MARKER AND SPACE
6030    IF LEFT$(TU$(K+1),1)<>"-" GOTO 6100
6035    IF LEN(TU$(K))<15 THEN
        TW$=TU$(K)+SPACE$(15-LEN(TU$(K)))
6040    GOSUB 6200
6042    TU$(K)=TU$(K)+" "   'ADD BACK
        SPACE
6045    IF TW$<>B.PRS$ THEN PRINT #5,
        TU$(K);" NOT IN VERB FILE.":GOTO
        6105
6050    ONS$=LEFT$(TU$(K+1),2)
6055    IF ONS$="-a" OR ONS$="-u" THEN
        STM$=B.INF$+" ":TU$(K)
        =LEFT$(STM$,INSTR(STM$," "))
6060    IF ONS$="-t" THEN STM$=B.PST$+"
        ":TU$(K)  =LEFT$(STM$,INSTR(STM$,"
        ")):TU$(K+1)="-"
        +RIGHT$(TU$(K+1),LEN(TU$(K+1))-2)
6065    IF ONS$="-i" THEN GOSUB 6300
6070    IF ONS$="-p" THEN STM$=B.FUT$+"
        ":TU$(K)  =LEFT$(STM$,INSTR(STM$,"
        ")):TU$(K+1)="-"  +RIGHT $(TU$(K+1),
        LEN(TU$(K+1))-2)
6072    WEND
6074 '  ========= NOUN SECTION =========
6075    WHILE
        RIGHT$(TU$(K),3)="N"+CHR$(2)+" "
6077    TU$(K)=LEFT$(TU$(K),LEN(TU$(K))-4)
        'CUT OFF MARKER AND SPACE
6079    IF LEFT$(TU$(K+1),1)<>"-" GOTO 6100
6081    WHILE RIGHT$(TU$(K),2)="am"
6083    TU$(K)=LEFT$(TU$(K),LEN(TU$(K))-
        2)+"att "
6085    WEND
6087    WHILE RIGHT$(TU$(K),2)="Du"
6089    TU$(K)=LEFT$(TU$(K),LEN(TU$(K))-
        2)+"TT "
6091    WEND
6093    WHILE RIGHT$(TU$(K),1)="e"
6095    TU$(K)=LEFT$(TU$(K),LEN(TU$(K))-
        1)+"ay "
6097    WEND
6100    IF RIGHT$(TU$(K),1)<>" " THEN
        TU$(K)=TU$(K)+" "'ADD SPACE IF NEC
6105    WEND
6110    K=K+1
6120    WEND
6130    RETURN
6140 '
6199    '*** SUBROUTINE FOR BINARY
        SEARCH OF VERB FILE ***
6200    UL=VSIZ          'normally size+1, but
        here last elt is dummy
6210    LL=1
6220    MDL=INT((UL+LL)/2)
6230    GET#3,MDL
6240    WHILE TW$<>B.PRS$ AND MDL<>LL
6250    IF TW$<B.PRS$ THEN UL=MDL ELSE
        LL=MDL
6260    MDL=INT((UL+LL)/2)
6270    GET#3,MDL
6280    WEND
6290    RETURN
6295 '
6299    '*** SUBROUTINE FOR PAST
        PARTICIPLE ***
6300    STM$=B.PPL$+" "
6305    TU$(K)=LEFT$(STM$,INSTR(STM$," "))
6310    IF LEFT$(TU$(K+1),7)<>"-irukk-" THEN
        MID$(TU$(K+1),2,1)
        =MID$(STM$,INSTR(STM$," ")-
        1,1):TU$(K)
        =LEFT$(STM$,INSTR(STM$," ")-2)+" "
6320    RETURN
6998 '
6999    '**** WORD CAPTURE SUBROUTINE
        ***
7000    W$=""
7005    WHILE LEFT$(L$,1)<>" " AND RM>0
7006    REM IGNORE PUNCTUATION
7007    F$=LEFT$(L$,1)
7008    IF F$="." OR F$="," OR F$="?" OR
        F$=":" OR F$=";" OR F$="|" GOTO 7020
7009    IF F$="!" OR F$="(" OR F$=")" GOTO
        7020
7010    W$=W$+F$
7012    IF F$=CHR$(21) THEN IC=IC+1
        'INCREMENT ITALIC CTRL CHAR
        COUNTER
7020    RM=RM-1
7025    L$=RIGHT$(L$,RM)
7030    WEND
7050    RETURN
7060 '
7500 '  *** ELIMINATION OF BRACKETED
        MATERIAL IN SOURASHTRA  INPUT ***
7560    REM INITIALIZE COUNTERS 7565  IO=0
7570    K=0
7590    WHILE K<I
7595    K=K+1
7600    IF LEFT$(SW$(K),1)<>"[" GOTO 7650
7610    BR$="]"
7620    GOSUB 7730
7630    GOTO 7590
7650    IF LEFT$(SW$(K),1)<>"{" GOTO 7685
7660    BR$="}"
7670    GOSUB 7730
7680    GOTO 7590
7685    IF LEFT$(SW$(K),1)<>"<" GOTO 7700
7686    BR$=">"
```

```
7687     GOSUB 7730
7688     GOTO 7590
7700    IF RIGHT$(SW$(K),1)="#" GOTO 7590
7702    IF RIGHT$(SW$(K),1)="]" OR
        RIGHT$(SW$(K),1)="}" OR
        RIGHT$(SW$(K),1)=">" THEN PRINT#5,
        "UNPAIRED ";RIGHT$(SW$(K),1);" IN
        LINE ";LI:GOTO 1300
7710    IO=IO+1
7715    SO$(IO)=SW$(K)
7720     WEND
7725     GOTO 7750
7728 '
7729 '  *** SUBROUTINE TO ELIMINATE
        BRACKETTED MATERIAL ***
7730    WHILE RIGHT$(SW$(K),1)<>BR$
7732     K=K+1
7734     IF K>I THEN PRINT#5,"MISSING
        ";BR$;" IN LINE ";LI:GOTO 1300
7736    WEND
7738     RETURN
7740 '
7750     ' *** MATCH SO$() ARRAY WITH ITEMS
        FROM GLOSSARY ***
7760    FOR K=1 TO IO
7764     IF LEN(SO$(K))>20 GOTO 7850
7765     SO$(K)=SO$(K)+SPACE$(20-
        LEN(SO$(K)))
7770     GOSUB 8000
7780    REM CHECK FOR NO MATCH
7790     IF SO$(K)=B.SGW$ GOTO 7800
7792    TU$(K)=SPACE$(10)
7793    IF LEFT$(SO$(K),1)=CHR$(21) THEN
        TU$(K)=LEFT$(SO$(K),INSTR(SO$(K),"
        "))
7795     GOTO 7850 7800          REM MATCH
        FOUND!
7810
        TU$(K)=LEFT$(B.TGW$,INSTR(B.TGW$,
        " "))
7850    NEXT K
7860    GOSUB 6000   'RUN MORPHOLOGICAL
        PROCESSOR
7870    GOSUB 5000   'RUN SANDHI
        PROCESSOR
7880    GOTO 8200
7998 '
7999     ' *** SUBROUTINE FOR BINARY
        SEARCH OF GLOSSARY ***
8000    UL=GSIZ+1
8010    LL=1
8020    MDL=INT((UL+LL)/2)
8030    GET#2,MDL
8040    WHILE SO$(K)<>B.SGW$ AND MDL<>LL
8050     IF SO$(K)<B.SGW$ THEN UL=MDL
        ELSE LL=MDL
8080     MDL=INT((UL+LL)/2)
8090     GET#2,MDL
8100    WEND
8120     RETURN
8199 '  *** OUTPUT A TURN ***
8200    K=1   'INITIALIZE COUNTER
8205    OLK=0
8210    WHILE K<=IO
8211     L$=""
8212     IC=0
8213     OLK=OLK+1
8215     IF OLK>ILK THEN
        LO$(OLK)=SPACE$(46) 'LO$(ILK) = line
        before COL BREAK
8218     IF LEFT$(TU$(K),1)=CHR$(21) THEN
        IC=IC+2
8220    IF LEN(L$)+LEN(TU$(K))>31+IC OR K>IO
        GOTO 8260
8230    L$=L$+TU$(K)
8240    K=K+1
8250     GOTO 8218
8260     PRINT#4,LO$(OLK)+L$
8270    WEND
8280    FOR K=OLK+1 TO ILK   'PRINT ANY
        REMAINING INPUT LINES
8290     PRINT#4,LO$(K)
8300    NEXT K
8310    PRINT#4,CHR$(11);CHR$(15)   'PRINT
        COLUMN BREAK
8320    GOTO 1351
8330 '
9000    CLOSE
9010    PRINT "OPERATION SUCCESSFULLY
        TERMINATED"
9011    BEEP:BEEP:BEEP:BEEP:BEEP
9020    END
```

3. EQUIVALENCE-PAIR EXTRACTION PROGRAM

```
10      'STD.BAS
12      'IAN SMITH, DLLL, YORK UNIVERSITY,
        NORTH YORK, ONT. CANADA M3J 1P3
14      'THIS VERSION 3 NOV 1989
16      'PROGRAMME INPUTS PARALLEL
        SOURASHTRA AND TAMIL TEXTS
17      'PREPARED WITH PC-WRITE (tm). (SEE
        FORMAT BELOW)
18      'AND OUTPUTS A LIST OF PAIRED
        WORDS. MATERIAL BETWEEN THE
        FOLLOWING
20      'IS IGNORED: <> {} [] (). SO IS
        PUNCTUATION.
21      'OUTPUT OCCURS WHENEVER A
        COLUMN-BREAK '(CHR$(11)+CHR$(15))
        IS FOUND.
22      'CONTROL LINES (OTHER THAN COL
        BRK) BEGINNING WITH CHR$(11)
23      'OR CHR$(12) (PG-BREAK) ARE
        IGNORED.
30      'FORMAT OF INPUT: SOU IN COLS 8-40;
        TAM IN 42+, IN DETERMINING THESE
32      'MARGINS, THE PROGRAMME IGNORES
        ITALIC CONTROL CHARACTERS
49      '
50      'FILES
52      '
54      ' F$   INPUT FILE SPECIFIED BY USER
```

```
56      'STD.OUT   OUTPUT FILE OF PAIRS OF
        MATCHED WORDS
57      'STD.ERR   ERROR FILE, CONTAINS
        NOTES ABOUT MISSING BRACKETS,
58      'UNEQUAL NUMBER OF SOU AND TAM
        WORDS IN A TURN
59      '
60      'VARIABLES
61      '
62      'SW$   ARRAY OF TENTATIVE
        SOURASHTRA WORD FORMS IN A TURN
64      'FROM INPUT FILE
66      'SO$ARRAY OF SOURASHTRA WORD
        FORMS IN A TURN AFTER
68      'WINNOWING
72      'TW$ARRAY OF TENTATIVE TAMIL
        WORD FORMS IN A TURN
74      ' FROM INPUT FILE
76      'TU$ARRAY OF TAMIL WORD FORMS
        IN A TURN AFTER
78      'WINNOWING
80      'BR$   TYPE OF BRACKET
82      'I   COUNTER FOR SW$ IN A TURN
84      'J   COUNTER FOR TW$ IN A TURN
86      'IO   COUNTER FOR SO$ IN A TURN
88      'JO   COUNTER FOR TU$ IN A TURN
90      'W$   TENTATIVE WORD CAPTURED BY
        SUBROUTING
92      'L$   LINE INPUT FROM INFPUT FILE /
        REMAINS OF SUCH LINE
93      'AFTER WORD HAS BEEN TRUNCATED
94      'LN   INITIAL LENGTH OF INPUT LINE
96      'LM   REMAINING LENGTH OF INPUT
        LINE
98      '
100   ' ********** PROGRAMME ************
101     DIM SW$(500)
102     DIM SO$(500)
103     DIM TW$(500)
104     DIM TU$(500)
105      REM initialize counters and flags
110     LI=0
200      INPUT "NAME OF FILE TO
        PROCESS",F$
210     OPEN "I",1,F$
220      REM  INPUT "NAME OF OUTPUT
        FILE",F$
221     F$="STD.OUT
230     OPEN "O",2,F$
235     OPEN "O",3,"STD.ERR"
299      REM START NEW TURN
300      REM RESET COUNTERS
351     I=0
352     J=0
400      IF EOF(1) GOTO 9000
402     IC=0      'SET COUNTER FOR ITALICS
405      LINE INPUT#1,L$
410     REM CHECK FOR COL BRK
415     IF LEFT$(L$,2)=CHR$(11)+CHR$(15)
        GOTO 7565
416     IF LEFT$(L$,1)=CHR$(12) GOTO 400
        'IGNORE PG BREAK
417     IF LEFT$(L$,1)=CHR$(11) GOTO 351
        'IF CONTROL LN, RESTART TURN
420      LI=LI+1
430      LN=LEN(L$)
435      IF LN<8 GOTO 400
540       REM RESET LENGTH AND CHECK FOR
        EOL
550      RM=LEN(L$)
560      IF RM<2 GOTO 400   'IF 1, MUST BE
        SPACE
570       REM REMOVE LEADING SPACES 575
        IF LEFT$(L$,1)=" " THEN
        L$=RIGHT$(L$,RM-1): GOTO 550
600       REM CAPTURE WORD
610      GOSUB 7000
700       REM CHECK FOR NULL WORD
710      IF W$="" GOTO 540
720       REM  ASSIGN TO SW OR TW ARRAY
725      IF LN-RM>45+IC THEN
        J=J+1:TW$(J)=W$:GOTO 550
730      I=I+1:SW$(I)=W$:GOTO 550
6998  '
6999      '*** WORD CAPTURE SUBROUTINE
        ***
7000       W$=""
7005       WHILE LEFT$(L$,1)<>" " AND RM>0
7006     REM IGNORE PUNCTUATION
7007       F$=LEFT$(L$,1)
7008       IF F$="." OR F$="," OR F$="?" OR
        F$=":" OR F$=";" OR F$="|" GOTO 7020
7009       IF F$="!" OR F$="(" OR F$=")" GOTO
        7020]
7010     W$=W$+F$

7011     IF F$=CHR$(21) THEN IC=IC+1
7020     RM=RM-1
7025     L$=RIGHT$(L$,RM)
7030      WEND
7050      RETURN
7499  '
7500      '*** ELIMINATION OF BRACKETED
        MATERIAL AND PREPARATION OF
        OUTPUT ***
7550      REM SOURASHTRA SECTION
7560      REM INITIALIZE COUNTERS
7565      IO=0
7570      K=0
7590      WHILE K<I
7595      K=K+1
7600     IF LEFT$(SW$(K),1)<>"[" GOTO 7650
7610      BR$="]"
7620      GOSUB 7730
7630      GOTO 7590
7650     IF LEFT$(SW$(K),1)<>"{" GOTO 7685
7660      BR$="}"
7670      GOSUB 7730
7680      GOTO 7590
7685     IF LEFT$(SW$(K),1)<>"<" GOTO 7700
7686      BR$=">"
7687      GOSUB 7730
7688      GOTO 7590
7700     IF RIGHT$(SW$(K),1)="#" GOTO 7590
7710     IO=IO+1
7715     SO$(IO)=SW$(K)
7720      WEND
7725      GOTO 7750
```

```
7728 '
7729    '*** SUBROUTINE TO ELIMINATE
        BRACKETED MATERIAL ***
7730     WHILE RIGHT$(SW$(K),1)<>BR$
7732    K=K+1
7734    IF K>I THEN PRINT#3,"MISSING ";BR$;"
        IN LINE ";LI:GOTO 300
7736     WEND
7738     RETURN
7749    '=========================
7750     REM TAMIL SECTION
7760     REM INITIALIZE COUNTERS
7765     JO=0
7770     K=0
7790     WHILE K<J
7795    K=K+1
7800    IF LEFT$(TW$(K),1)<>"[" GOTO 7850
7810     BR$="]"
7820     GOSUB 7930
7830     GOTO 7790
7850    IF LEFT$(TW$(K),1)<>"{" GOTO 7900
7860     BR$="}"
7870     GOSUB 7930
7880     GOTO 7790
7900    IF RIGHT$(TW$(K),1)="#" GOTO 7790
7910    JO=JO+1
7915    TU$(JO)=TW$(K)
7920     WEND
7925     GOTO 8000
7928 '
7929    '*** SUBROUTINE TO ELIMINATE
        BRACKETED MATERIAL ***
7930    WHILE RIGHT$(TW$(K),1)<>BR$
7932     K=K+1
7934     IF K>J THEN PRINT#3, "MISSING
        ";BR$;" IN LINE ";LI:GOTO 300
7936    WEND
7938     RETURN
7998 '
7999    '************ PRINTOUT SECTION
        **************
8000    REM CHECK FOR EQUAL NO OF TW AND
        SW
8015     IF IO=JO GOTO 8050 8020
                REM ERROR SECTION
8030     PRINT#3,"WORD COUNT MISMATCH
        IN LINE ";LI
8035     FOR K=1 TO IO:PRINT#3, SO$(K);"
        ";:NEXT K
8040     FOR K=1 TO JO:PRINT#3, TU$(K);"
        ";:NEXT K
8045     GOTO 300
8050    REM WRITE SW AND TW PAIRS TO
        OUTPUT FILE
8060    FOR K=1 TO IO:PRINT#2,
        SO$(K);",",TU$(K):NEXT K
8070    GOTO 300
9000     CLOSE
9010     PRINT "OPERATION SUCCESSFULLY
        TERMINATED"
9011    BEEP:BEEP:BEEP
9020     END
```

4. GLOSSARY COMPARISON PROGRAM

```
50 '    GCOMP.BAS
60 '
65 '    FIRST CREATED July 6, 1990  IAN
        SMITH, YORK UNIVERSITY
70 '    LAST UPDATE July 9, 1990  IAN SMITH,
        YORK UNIVERSITY
80 '
90 '    REPLACES EARLIER GENERATION OF
        GCOMP.BAS (NOV '89 - JUNE '90)
95 '
100 '   PROGRAMME TO COMPARE LIST OF
        PAIRED VOCABULARY ITEMS IN
        STD.OUT WITH
110 '   ITEMS IN EXISTING GLOSSARY.
        STD.OUT IS FIRST CONVERTED TO A
        RANDOM FILE
115 '   AND SORTED; OVERLENGTH ITEMS ARE
        SENT TO GCOMP.ERR. COMPARISON
        WITH
117 '   GLOSSARY IS THEN MADE. DUPLICATE
        ITEMS ARE ELIMINATED AT THIS
        STAGE.
118 '   OUTPUT FILE "GCOMP.SRT" CONTAINS
        ITEMS NOT FOUND. FILE
        "GCOMP.QRY"
119 '   CONTAINS ITEMS FOR WHICH THE
        SOURASHTRA WORD MATCHED BUT
        FOR WHICH THE
121 '   TAMIL WORD DID NOT. NOTE: THIS
        VERSION REQUIRES RAMDISK D:
122 '   WHICH IS NOT THE DEFAULT.
123 '
140 '   FILES 150 '
160 '   GLOSSARY.ST  EXISTING GLOSSARY
        (RANDOM-ACCESS)
162 '   - 2 FIELDS, BUFFERS: B.SW$ (20),
        B.TW$ (20) USED IN
163 '   CONVERSION TO RANDOM AND
        COMPARISON WITH EXISTING
        GLOSSARY
170 '   GCOMP.SRT  OUTPUT FILE
        (SEQUENTIAL, FORMATTED WITH
        PADDING) SEE ABOVE
175 '   GCOMP.QRY      OUTPUT FILE
        (SEQUENTIAL, FORMATTED WITH
        PADDING) SEE ABOVE
180 '   STD.OUT  OUTPUT OF STD.BAS
        (SEQUENTIAL) - DIFFERENT NAME CAN
        BE
181 '   INPUT FROM KEYBOARD.
182 '   STD.RND  RANDOM VERSION OF
        STD.OUT
183 '     - 1 FIELD, BUFFERS: B.L$ (40)  USED
        IN SORTING ROUTINE
190 '   GLOSSARY.DAT  FILE CONTAINING
        GLOSSARY DATA: NO OF ENTRIES
200 '   GCOMP.ERR  ERROR FILE
210 '
220 '   G  GAP FOR SHELL SORT
```

```
230 '   I  COUNTER FOR NO OF LINES IN FILE
        TO SORT
240 '   SF  SWAP FLAG: 1 = NO SWAPS MADE;
        0 = SWAP MADE THIS PASS
250 '   PRV$  PREVIOUS PAIR FROM STD.RND
260 '   SW$  SOURASHTRA WORD FROM
        STD.OUT (OR F$)
270 '   TW$  TAMIL WORD FROM STD.OUT
        (OR F$)
280 '   GSIZ  NO OF ENTRIES IN GLOSSARY.ST
290 '   STDSIZ  NO OF ENTRIRES IN STD.RND
318 '   F$  FILENAME INPUT FROM
        KEYBOARD
320 '   B.SGW$  SOURASHTRA WORD BUFFER
        FOR GLOSSARY.ST
322 '   B.TGW$  TAMIL WORD BUFFER FOR
        GLOSSARY.ST
324 '   B.SW$  SOURASHTRA WORD BUFFER
        FOR STD.RND
326 '   B.TW$  TAMIL WORD BUFFER FOR
        STD.RND
399 '
400 '   *** PROGRAMME MAINLINE ***
401     CLS
402     LOCATE 3,20
403     PRINT "GCOMP - GLOSSARY
        COMPARISON PROGRAM"
405     OPEN "O",5,"GCOMP.ERR"
410     GOSUB 999   'CONVERT STD.OUT
        STD.RND
420     GOSUB 1499   'SORT STD.RND
430     GOSUB 1700   'COMPARE STD.RND TO
        GLOSSARY
450     END
998 '
999 '   *** SUBROUTINE TO CONVERT
        SEQUENTIAL STD.OUT TO RANDOM ***
1000    LOCATE 5,5:PRINT "CONVERTING
        INPUT FILE TO RANDOM ON RAMDISK
        D:"
1048    LOCATE 6,5:INPUT "NAME OF
        SEQUENTIAL INPUT FILE
        (DEFAULT=STD.OUT): ",F$
1049    IF F$="" THEN F$="STD.OUT"
1050    OPEN "I", 1, F$
1060    OPEN "R", 2, "D:STD.RND", 40
1070    FIELD #2, 20 AS B.SW$, 20 AS B.TW$
1100    LOCATE 7,5:PRINT "CONVERTING ..... "
1105    WHILE NOT EOF(1)
1109     REM INPUT A PAIR FROM STD.OUT,
        CHECK LENGTH OF SW$/TW$
1110    INPUT #1,SW$,TW$
1111    IF LEN(TW$)>20 OR
        LEN(SW$)>20...THEN
        PRINT#5,"OVERLENGTH ITEM:
        ",SW$,TW$:GOTO 1110
1119     REM LENGTH OK - OUTPUT PAIR TO
        STD.RND
1120     LSET B.SW$=SW$
1130     LSET B.TW$=TW$
1140     PUT#2
1150    WEND
1160    STDSIZ=LOC(2)
1170    LSET B.SW$=CHR$(26)  'ONLY TO
        MAKE THE DOS "TYPE" COMMAND
        WORK
1180    PUT#2
1300    CLOSE #1,2
1335    LOCATE 9,5
1336    PRINT "CONVERSION SUCCESSFULLY
        COMPLETED."
1337    LOCATE 10,10
1338    PRINT "NO. OF ENTRIES = ";STDSIZ
1400    RETURN
1410 '
1499 '  *** SHELLSORT OF STD.RND ***
1500    LOCATE 12,5:PRINT "SORTING
        STD.RND"
1509    OPEN "R",2,"D:STD.RND",40
1510    FIELD #2, 40 AS B.L$
1515    J=STDSIZ
1520    G=INT(J/2)
1522    LOCATE 13,10:PRINT "SHELL SIZE =
        ";G;"  PASS NO. "
1525    WHILE G>0
1530     SF=0:PN=0
1540     WHILE SF=0
1550     SF= 1:PN=PN+1
1555     LOCATE 13,39:PRINT USING "###";PN
1560     FOR K=1 TO J-G
1562    GET#2,K 1564   L1$=B.L$
1566    GET#2,K+G
1568    L2$=B.L$
1570    IF L1$ > L2$ THEN PUT#2,K:LSET
        B.L$=L1$:PUT#2, K+G:SF=0
1580     NEXT K
1590    WEND
1600    G=INT(G/2)
1602    LOCATE 13,23:PRINT G
1610    WEND
1620    LOCATE 14,5:PRINT "SORT COMPLETE"
1630    CLOSE #2
1640    RETURN
1698 '
1699    '*** COMPARISON OF STD.RND WITH
        GLOSSARY.ST ***
1700    LOCATE 16,5:PRINT "BEGINNING
        COMPARISON OF STD.RND WITH
        GLOSSARY.ST
1719     REM OPEN FILES FOR INPUT AND
        OUTPUT
1720    OPEN "R",2,"GLOSSARY.ST",40
1730    FIELD #2, 20 AS B.SGW$,20 AS B.TGW$
1740    OPEN "I",3,"GLOSSARY.DAT"
1750     REM INPUT GLOSSARY SIZE
1760    INPUT#3, GSIZ
1770     CLOSE #3
1775     ' ***** section to be reinstated if big
        ramdisk available ****
1780     'OPEN "R",3,"GLOSSARY.ST",40
1790     'FIELD #3, 20 AS B.SW$,20 AS B.TW$
1800     'PRINT "TRANSFERRING GLOSSARY TO
        RAMDISK ..."
1810     'FOR I=1 TO GSIZ
1820 '  GET#3
1830'   LSET B.SGW$=B.SW$
1840 '  LSET B.TGW$=B.TW$
```

```
1850 '   PUT#2
1860      'NEXT I
1870      'CLOSE #3
1880      'PRINT "GLOSSARY SUCCESSFULLY
          TRANSFERRED.  NO. OF ENTRIES =
          ";GSIZ
1890      OPEN "R",1,"D:STD.RND",40
1900      FIELD #1, 20 AS B.SW$,20 AS B.TW$
1910       REM INITIALIZE COUNTERS
1920       NM=0
1930       QY=0
1935       PRV$=""
1937       IP=0:IC=0
1940       REM OPEN FURTHER OUTPUT FILES
1950  OPEN "O",3,"GCOMP.SRT"
1960       OPEN "O",4,"GCOMP.QRY"
1970      PRINT#4,"FIRST ITEM OF EACH PAIR IS
          FROM STD.OUT, SECOND FROM
          GLOSSARY"
1980      REM INPUT A PAIR FROM
          GLOSSARY.ST
1990       GET#2
2000      WHILE LOC(1)<STDSIZ
2011       GET#1:IP=IP+1:IC=IC+1   'INPUT FROM
          STD.RND
2012       IF IC>=100  THEN LOCATE 17,10:PRINT
          IP;" OF ";STDSIZ;" ITEMS
          PROCESSED":IC=0
2015      IF B.SW$+B.TW$=PRV$ THEN 2011
2017      PRV$=B.SW$+B.TW$
2020      WHILE B.SGW$<B.SW$ AND
          LOC(2)<GSIZ
2030       GET#2
2035      WEND
2040      REM CHECK FOR NO MATCH
2045      IF B.SW$<>B.SGW$ THEN GOSUB
          8200:GOTO 2079
2050      REM IF MATCH FOR SW$, CHECK THAT
          TW$ IS SAME TOO
2055      IF B.TW$=B.TGW$ THEN 2079
2056      REM REMOVE CATEGORY MARKING, IF
          ANY FROM TGW$ AND CHECK FOR
          IDENTITY
2057
          TGW$=B.TGW$:CM=INSTR(TGW$,CHR$
          (2))
2058       IF CM>0 THEN MID$(TGW$,CM)="  "
2060       IF B.TW$=TGW$ THEN 2079
2062      REM NOW SEE IF IT IS AN AFFIX:
          APOLOGIES FOR THE SPAGHETTI
2064       IF LEFT$(B.TW$,1)<>"-" OR
          LEFT$(B.TGW$,1)<>"-" THEN 2078
2066       SP=INSTR(B.TGW$," ")
2067       IF SP>0 THEN
          TGW$=LEFT$(B.TGW$,SP-1)
2068       IF INSTR(B.TW$,TGW$)=1 THEN 2079
2069       TW$="-i"+RIGHT$(B.TW$,18)
2072       IF INSTR(TW$,TGW$)=1 THEN 2079
2078      GOSUB 8300
2079        WEND
2080       CLOSE
2090       LOCATE 19,5
2100       PRINT "GLOSSARY COMPARISON
          SUCCESSFULLY TERMINATED"
2105       LOCATE 20,5
2110       PRINT "NO. OF ITEMS FOR WHICH NO
          MATCH FOUND: ";NM
2115       LOCATE 21,52120 PRINT "NO. OF
          ITEMS FOR WHICH PARTIAL MATCH
          FOUND: ";QY
2130      BEEP:BEEP:BEEP
2210       RETURN
8198 '
8199      '***** OUTPUT NEW ITEMS TO
          GCOMP.SRT *****
8200      PRINT#3,B.SW$;",";B.TW$
8210      NM=NM+1
8230       RETURN
8298 '
8299      '**** OUTPUT PARTIAL MATCH ITEMS
          TO GCOMP.QRY ****
8300      PRINT#4,
          B.SW$;B.TW$:PRINT#4,B.SGW$;
          B.TGW$: PRINT#4, CHR$(13)+CHR$(10)
8310      QY=QY+1
8320       RETURN
9000 '   *** END OF PROGRAMME ***
```

MEN'S AND WOMEN'S SPEECH IN JAPANESE

László Szabó
University of New Brunswick

ABSTRACT

In Japanese there are several politeness levels, such as superpolite (honorific or humble), formal and informal. In polite and formal speech, women and men use more or less the same expressions. But in informal, abrupt speech, both men and women have their own style. This is the topic of this paper. There are men's and women's words, men's or women's particles, grammatical forms and constructions.

In the Japanese language, there are many ways to express formality or politeness. The Japanese can use superpolite words and expressions if they speak to a highly honorable person. The superpoliteness can be honorific if there is reference to the addressed honorable person. It can be humble if the speaker refers to himself, expressing his respect towards somebody by referring to his own modest, humble existence. The formal style is the most common way of speaking, which is not superpolite but polite enough to be used in most situations. In all these polite and formal expressions, men and women use more or less the same words and constructions. If the formal verb (ending in the suffix complexes *-masu*, *-masen*, *-mashita*, *-mashoo* or containing the copula *desu*, *deshita*, *deshoo*, etc.) is used, the style is formal. In these cases there is no difference between men's and women's speech.

However, the Japanese like to use informal, abrupt expressions, both men and women. In informal style, men's and women's speech differs significantly. This is the topic of the present paper.

There are many words for 'yes' and 'no'. In formal speech, the most common word for 'yes' is *hai*. Women use *ee* in informal speech, while men say *un* or *aa*. I quote examples:

> Formal: *Hai, soo desu* 'Yes, it is so'; women's informal: *Ee, soo yo* [idem] men's informal: *Un, soo da yo* [idem].

> Formal: *Hai, yatto dekimashita* 'Yes, it (i.e. a house) was finally completed'; women's informal: *Ee,yatto dekita* [idem]:men's informal: *Aa. yatto dekita* [idem].

The word for 'no' is *iie* in formal speech. In informal speech women say *uun*, men say *iya*. Examples:

> Formal: *Iie, (watakushi-wa) o-sake-wa nominasen* 'No, I don't drink sake'; women's informal: *Uun, o-sake wa nomanai-wa* [idem] men's informal: *Iya, sake-wa nomanai* yo [idem].

> Formal: *Iie, amari suki ja arimasen* 'No, I don't like it very much'; women's informal: *Uun, ammari suki ja nai-wa* [idem]: men's informal: *Iya, amari suki ja nai n-da* [idem].

The Japanese do not use the words for 'I', 'you', 'he, she', etc., very often, but they have more words for them than most other languages, enough to distinguish between formal, women's informal and men's informal speech.

The word for 'I' is *watakushi* or *watashi* in formal speech. Women say informally *atashi* 'I'. Men usually use the informal men's word *boku* 'I'; the other informal men's word for 'I' is *ore* which is not used as often. Examples:

> Formal: *Watakushi-wa hitori-de sunde iru node* 'Because I live alone'; women's informal: *Atashi hitori-de sunderu-n-de* [idem]: men's informal: *Boku hitori-de sunderu-n-de* [idem].

> Formal: *Watakushi-mo anata-o o-sasoi shiyoo to omotte denwa-o shimashita* 'I, too, phoned with the intention of asking you out'; women's informal: *Atashi-mo anata-o sasoi dasoo to omotte denwa shita no yo* [idem] men's informal: *Ore mo kimi-o sasoi dasoo to omotte denwa shita n'da yo* [idem].

The word for 'you' is *anata* in formal style as well as women's informal speech. Men use informally the word *kimi* 'you'. Examples:

> Formal: *Anata-no uchi-wa doko desu ka?* 'Where is your house?'; women's informal: *Anata-no uchi doko?* [idem] men's informal: *Kimi-no uchi doko?* [idem].

All words for 'I', 'you', 'he, she', etc., behave grammatically as nouns. This is the reason why they are not taken for personal pronouns. This is still more evident in the case of the words for 'he' and 'she'. There is no masculine or feminine gender in this language, but under the influence of translations into Japanese, the use of the word *kare* 'he' and *kanojo* 'she' is getting more and more common. However, many people prefer using the construction *ano kata* that person' instead of the equivalents of 'he' or 'she'. This occurs particularly in formal or women's informal speech, while in men's informal speech we can hear *kare* 'he', etc. Examples:

> Formal: *Ano kata-wa sensei deshita ka?* 'Was he a teacher?' women's informal: *Ano kata-wa sensei datta no?* [idem] men's informal: *Kare sensei datta kai?* [idem].

The speakers of this language do not distinguish between Mr., Mrs., Miss, etc. They use the suffix *-san* which can join names, words for professions, etc. One could take this suffix for the equivalent of all these English words: Mr., Mrs., Miss, etc. This suffix is used in formal as well as women's informal speech. Men use informally the suffix *-kun* which makes the address form very friendly, something like an older brother talking to the younger brother. Examples:

> Formal: *Howaito-san-wa kao-iro-ga warui-wa ne* 'You, Mr. White, don't look well'. This sentence is exactly the same in women's informal speech, while men say informally: *Howaito-kun kao-iro-ga warui-ne* [idem].

Talking about 'my own mother', they use the noun *haha* both in formal and in women's informal speech. Men can use informally the noun *ofukuro* '(my) mother'. Examples:

> Formal: *Haha-ni nani-ka gochisoo-o tsukutte moraimasu a* 'I shall have my mother prepare something good ; women's informal: *Haha-ni nani-ka g isoo tsukutte morau kara* [idem]; men's informal: *Ofukuro-ni nani-ka gochisoo tsukutte mora kara* [idem].

There are many more words which are used either in men's or in women's informal style. There is no space for all of them in this paper. I will just mention a few more without quoting sentences. Men say *kure* as the equivalent of the polite word *kudasai* 'please', 'do it for me', while women say *choodai*. Instead of the formal expression *onegai shimasu* 'please do it', men like to say *tanomu yo* [idem] Even such adverbs as *taihen* 'very' can be different in men's and women's speech. Women prefer the adverb *totemo*, while in men's informal speech the adverb *sogoku* [idem] is used.

The use of certain particles at the end of the sentence is very characteristic and makes a very noticeable difference between women's and men's informal speech. I present examples for the most important particles.

The final particle *no* makes the sentence in informal women's speech milder, more 'feminine'. Examples:

> Formal: *Suehiro-de bifuteki-o tabemashita* 'I ate beef steak at Suehiro'; women's informal: *Suehiro-de bifuteki-o tabeta no* [idem] men's informal: *Suehiro-de bifuteki-o tabeta n-da* [idem].

Men's informal sentences might contain the particle complex *n-dai* (in questions which require more than a 'yes' or 'no' answer.). Examples:

> Formal: *Dono depaato-ni hairimashita ka* 'Which department store did you enter?' women's informal: *Dono depaato-ni haitta no*? [idem] men's informal: *Dono depaato-ni haitta n-dai?* [idem].

The sentence final particle *wa* makes a woman's informal sentence milder, similar to the particle *no*. In the equivalent men's informal sentence we often find the particle *yo*. Examples:

> Formal: *Takusan arimasu* 'There are many (stores in a district)'; women's informal: *Takusan aru wa* [idem] men's informal: *Takusan aru yo* [idem].

Since this is very common and characteristic, I quote one more example:

> Formal: *Ginza-e ikimashita* 'I went to the Ginza'; women's informal: *Ginza-e itta wa* [idem] men's informal: *Ginza-e itta yo* [idem].

There are many more sentence final particles which make men's or women's speech different. I mention only one more: the particle *kai* is added to 'yes' or 'no' questions in men's informal speech, while in women's informal speech, in the equivalents of such questions there is no final particle. Examples:

> Formal: *Hontoo desu ka?* 'Is this true?'; women's informal: *Hontoo?* [idem] men's informal: *Hontoo kai?* [idem].

Here is one more example for this very common phenomenon:

> Formal: *Osake-mo nomimashita-ka?* 'Did you drink sake, too?'; women's informal: *Osake mo nonda* [idem] men's informal: *Sake mo nonda kai?* [idem].

In many expressions, when they use the formal existence verb *desu* in the sentence, there is no existence verb in women's informal style. However, in the equivalent sentence in men's informal speech we meet the abrupt equivalent of the existence verb, which is *da*. Examples:

Formal: *Ha ga itai-n desu* 'I have a toothache'; women's informal: *Ha ga itai no* [idem] men's informal: *Ha ga itai-n da* [idem].

I would like to mention here one more grammatical difference. The formal tentative (with the ending complex *-mashoo)* is characteristic in formal as well as women's informal style. In men's informal style they use the informal tentative (which ends in the suffix *-oo* or *-yoo*). Examples:

Formal and women's informal: *Ikimashoo* 'Let us go'; men's informal: *Ikoo* [idem].

Here is one more example:

Formal: *Dewa, dekakemashoo* 'Then, let us go'; women's informal: *Ja, dekakemashoo* [idem] men's informal: *Ja, dekakeyoo* [idem].

Without further detailed study and examples, I will also just mention that there are many more differences between men's and women's informal speech. Women's speech is milder, closer to formal speech, sometimes even to superpolite speech. My intention was to study here only the most characteristic and most common differences.

PRINTING
SERVICES MUN