

UC Santa Barbara

NCGIA Technical Reports

Title

Spatial Language and Geographic Information Systems: Cross-Linguistic Issues/ El Lenguaje Especial y Los Sistemas de Informacion Geograficos: Temas Interlinguisticos (90-2)

Permalink

<https://escholarship.org/uc/item/1tj51545>

Authors

Mark, David M.
Gould, Michael D.
Nunes, Joan

Publication Date

1990

**SPATIAL LANGUAGE AND GEOGRAPHIC INFORMATION
SYSTEMS: CROSS-LINGUISTIC ISSUES**

**(EL LENGUAJE ESPACIAL Y LOS SISTEMAS DE
INFORMACION GEOGRAFICOS: TEMAS
INTERLINGUISTICOS)**

by

David M. Mark¹, Michael D. Gould¹, and Joan Nunes²

National Center for Geographic Information and Analysis

Report 90-2

1. NCGIA, Department of Geography, University at Buffalo, Buffalo, NY. 14260
2. Departament de Geografia, Universitat Autònoma de Barcelona, 08193 Bellaterra, España

ABSTRACT

The great majority of existing geographic information systems have been designed by English or German speakers. Since human natural languages impose structure on the cognition and perception of space, time, and other concepts, GIS data models, and especially GIS query languages and human interfaces, can be expected to contain artifacts of the language spoken by their designers, most commonly English. At a practical level, natural language studies, particularly those based on cognitive linguistics, are important issues for the designers of geographical information systems, because GISs of the future should be able to handle natural language in a number of situations. We propose that the design of any natural language modules for GIS should be based on language universals and language primitives derived from cross-linguistic studies of geographic concepts and relations. Otherwise, it will be even more difficult to transfer English-designed systems into the many other languages that can be expected to make up an ever-increasing proportion of GIS markets in the future. This paper expands on the general principles of cognitive linguistics mentioned above, with emphasis on cross-linguistic issues. Then, we will present a more detailed study of the primitive geographic relations represented in most Indo-European languages by prepositions. We will concentrate on differences between English and Spanish, but also draw on other languages.

RESUMEN

La gran mayoría de los sistemas de información geográficos existentes han sido diseñados por hablantes de lengua inglesa o alemana. Como los lenguajes humanos naturales imponen una estructura en la cognición y percepción del espacio, tiempo y otros conceptos, cabe esperar que los modelos de datos y especialmente los lenguajes de consulta de los SIGs y el interface del usuario, contengan artefactos de la lengua hablada por sus diseñadores, el inglés en la mayoría de los casos. A nivel práctico los estudios sobre lenguajes naturales, particularmente esos basados en lingüística cognitiva, son temas importantes para los diseñadores de sistemas de información geográficos porque los SIGs del futuro deberían de ser capaces de manejar lenguajes naturales en diversas situaciones. El diseño de cualquier módulo de lenguaje natural para SIGs debería de estar basado en lenguas universales y lenguas primitivas derivadas de estudios interlingüísticos sobre conceptos y relaciones geográficos. De otro modo, será incluso más difícil transferir los sistemas diseñados en inglés a las muchas lenguas en las que se espera una proporción creciente de mercados para los SIGs en el futuro. Este trabajo desarrolla los principios generales de la lingüística cognitiva mencionada anteriormente, con énfasis en temas interlingüísticos. A continuación presentaremos un estudio más detallado de las relaciones geográficas primitivas representada por preposiciones en la mayoría de las lenguas europeas. Nos concentraremos en las diferencias entre el inglés y el español, pero también usaremos otras lenguas.

PREFACE AND ACKNOWLEDGEMENTS

This paper was originally written for the Second Latin American Conference on Applications of Geographic Information System Technology, held from September 25 to 29, 1989 in Mérida, Venezuela. Although the *Proceedings* of the Mérida conference have been distributed to the participants, that audience represents such a small minority of potentially interested GIS and other geographic professionals that we are reprinting the paper here. This "version" is essentially a copy of that in the Mérida proceedings, with only slight changes made to clarify some points.

The authors wish to thank Hsueh-chen Chou, Zuzana Dobes, Scott Freundsuh, and Victor Wu, who participated with the authors in a graduate seminar during which many of the ideas in this paper were developed. Freundsuh also provided useful comments on an earlier draft of this paper. We wish to thank Celso Alvarez, Jeff Lansing, Lucy Moran, Lior Moscovici, and Jack Tuszynski for valuable insights, examples, and references provided through electronic mail. Finally, thanks go to Dr. Andrew A. Frank for kindly presenting this paper in Mérida, after Hurricane Hugo changed our travel plans sufficiently so that we could not be at the conference.

This paper represents part of Research Initiative #2, "Languages of Spatial Relations", of the National Center for Geographic Information and Analysis, and was supported by a grant from the National Science Foundation (SES-88-10917). Support by NSF is gratefully acknowledged. Joan Nunes was able to visit the Buffalo site of the NCGIA thanks to a grant from the Dirección General de Investigación Científica y Técnica (DGICYT) of the Spanish Ministry of Education.

By electronic mail, the authors can be reached at geodmm@ubvms.bitnet [Mark], v010kkye@ubvms.bitnet [Gould], and ilge5@ebccuabl.bitnet [Nunes]

INTRODUCTION

Debido a que los principales grupos lingüísticos difieren en el modo de estructurar y expresar el espacio geográfico, el hecho de que los últimos sistemas de información geográficos han sido diseñados por hablantes de lengua inglesa o alemana es un impedimento para la adopción y uso efectivo de SIGs en países de lengua romance (como el español y el francés) y especialmente en países de lenguas no ,indo-europas.

Because major language groups differ in the ways they structure and express geographic space, the fact that most current geographic information systems have been designed by native speakers of English or German is a potential impediment to the widespread adoption and effective use of GIS in countries where people speak Romance languages (such as Spanish and French) and especially .non-Indo-European languages.

Whereas human senses operate in very similar ways, regardless of culture or language, human *perception* (that is, the *mental interpretation* of sensory inputs) is influenced by language and interpretive image schemas (see Lakoff, 1987, and Johnson, 1988). The title of Leonard Talmy's (1983) seminal paper on this subject, "How Language Structures Space", expresses this position very well. Perhaps more precisely, a human natural language reveals the cognitive categories that are most salient to its speakers, and these almost certainly provide a "top-down" structure for sensory input from the real world. Most aspects of perception and cognition are influenced in this way by language, but cognitive models of geographic space are of the highest interest to geographers and other professionals who deal with space and with objects and phenomena distributed within it.

From principles of cognitive science and human factors, we can predict that the ideal Geographic Information System (GIS) for an individual to use would define spatial objects and spatial relationships in a way that is consistent with that person's language, culture, and background. But this implies that such an "ideal" GIS would not be ideal for speakers of other languages, or even for speakers of the same language who belong to different cultures, dialects, or disciplines. Thus a cross -linguistic perspective on cognitive categories for geographic space raises serious issues for GIS technology transfer across cultural and especially linguistic boundaries.

The great majority of existing commercial and educational geographic information systems have been designed "in English" (or "in German"). We say that the design of a system occurs "in" a language because of the issues we raised in the previous section. Thus, GIS query languages, human interfaces, and even GIS data models (e.g. those of Arc/Info, object-oriented systems, etc.) may contain artifacts of the language spoken by their designers, most commonly English. Systems which mimic human cognitive structure and behavior are more likely to be useful and effective than systems which are designed based primarily upon current hardware and software trends. This may be especially true for systems which are to be used by novices or by non-geographic personnel, such as tourists or real estate (property) agents.

Even if we were not considering the cross-linguistic aspects. of language for geographic space, the role of natural language processing in GIS nevertheless is controversial. Some people in the field contend that interaction with a GIS should be non-verbal, and instead of words should use direct manipulation of command icons and geographic objects to the maximum degree possible. The Macintosh computer interface is a good example on the approach they recommend. However, we believe that natural language studies, particularly those based on cognitive linguistics, are important topics for the designers and users of GIS in several distinct areas:

- 1 . *Input of GIS queries and commands in natural language.* Natural language queries in typed form may be of only limited utility; however, when real-time understanding of normal speech

becomes practical, natural-language commands and queries may become a common form of system interaction.

2. *Input of geographic data and information in text form.* This includes current practical problems such as the computer processing of biological localities data (see McGranaghan, 1989). It might also extend to automated interpretation of explorer's journals, or of tape-recorded field notes.
3. *Natural language production for limited topic domains.* Generation of verbal descriptions of routes for drivers already is possible (Streeter et al., 1985; Mark, 1985; Ma, 1987; McGranaghan et al., 1987). Progress can be expected for other relatively-simple cases and domains, such as deed descriptions for land records.
4. *General natural language text generation.* The GIS of the future might produce grammatically-correct paragraphs describing locations, shapes, and patterns, for direct inclusion in reports or manuscripts.
5. *Establishing a GIS conceptual framework, including data structures, based on principles of human spatial cognition.* We feel it especially important to stress two points. Firstly, we are suggesting building data structures which mimic human cognitive structures (e.g. the "neighborhood"), and not the converse (such as by constraining human description of space through insistence on the use of "parcels", "polygons", etc., or of rectilinear grids, which may not support human cognitive structure but rather may circumvent such natural structure). Secondly, the process of building these data structures and thus representing all (or many) geographic spatial relations cannot be judged either possible or impossible until an "all-out", concerted effort is made. And, as Kuipers (1978) states, spatial knowledge is always a partial knowledge--it is never complete. Therefore, a partial representation of spatial knowledge in data structures is not wrong, but rather is simply incomplete and nonetheless may be useful. Concepts of space and spatial relations for GIS thus should be based on principles of human spatial cognition (Mark, Svorou, and Zubin, 1987; Mark, 1988, 1989; Mark and Frank, 1989; Mark et al., 1989). Studies of natural language represent a very important source for building and testing models of human spatial cognition.

In this paper, we focus on a sixth important natural- language issue for GIS: the cross-linguistic issues raised in our introductory paragraph, and the implications of these issues for GIS technology transfer across linguistic boundaries. Cross -linguistic studies of spatial language recently were identified as critical issues for the U.S. National Center for Geographic Information and Analysis (NCGIA) Research Initiative 2, "Languages of Spatial Relations" (Mark *et al.*, 1989; Mark, 1989), and Spanish was noted as being of especially high priority because it is one of the top 5 languages as far as the number of speakers worldwide.

The first section of the paper will review cognitive models of geographic space and spatial language. Next, we discuss issues involving human-computer interactions, focussing on cross-linguistic issues related to interface design, commands, and even translations of terms and manuals. Then, we present an informal discussion of some areas in which spatial representation in English and Spanish differ. Lastly, we present some priorities for future research.

SPATIAL LANGUAGE

Talmy (1983) discussed the cognitive basis for spatial language under the title "How language structures space." His work is based on the principle that much of the apparently "objective" structure of the real world is

imposed by our conceptual models and categories, and that these categories are determined, or at least revealed, by language. This is consistent with *experiential realism*, a model of categories and cognition advanced by Lakoff (1987) and Johnson (1988). Talmy points out that many locative expressions consist of a figure (object being located), and ground (or reference object), and a relation between them. In the sentence: "The cookies are in the jar", the cookies are the figure being located, the jar forms the ground or reference for specifying the location, and "in" represents their spatial relation.

In the Indo-European languages, such relations commonly are represented by prepositions, but some other languages use other grammatical constructions. Computation of an appropriate locative expression for a specified figural object involves first choosing an appropriate reference or ground object, then formally representing the nature of the ground and of the figure-ground relation, and finally encoding the relation linguistically.

Herskovits (1985, 1987) discussed locative expressions in English from a computational and cognitive viewpoint. Of particular interest is the chapter in her book (Herskovits, 1987) in which she discusses "use types" for the three "fundamental" spatial prepositions in English: "at", "on", and "in". The "ideal meaning" of "at" is: "for a point to be coincident with another point" (Herskovits, 1987, p. 128). Herskovits (p. 140) gives a more complicated ideal meaning for "on": "for a geometrical construct X to be contiguous with a line or surface Y; if Y is the surface of an object Oy, and X is the space occupied by another object Ox, for Oy to support Ox." Thirdly, the ideal meaning for "in" is: "inclusion of a geometric construct in a one-, two-, or three-dimensional geometric construct" (p. 149). Herskovits (1987) then listed 8 basic use types for "at", 11 for "on", and 11 for "in". (One of her use types for "on" is discussed in detail below.) One of our research objectives is to evaluate these thirty English-language use types from a cross-linguistic perspective.

Mark, Svorou, and Zubin (1987) discussed a number of cognitive and linguistic issues of relevance to GIS and to geographical analysis. In particular, they focussed on reference frames and on uncertain (fuzzy) spatial concepts such as "near". Depending on the situation, human languages may use gestalt reference frames, based on inherent properties of the ground or reference object, or canonical reference frames, based on the speakers' or listeners' viewpoint. There also are cross-linguistic and cross-cultural differences in reference frames; a good example is the common use of radial rather than Cartesian orientation systems by island-dwellers. Mark et al. also pointed out that scale- and context-dependency of the meaning of near may be more important than the fuzziness of that meaning.

Pullar and Egenhofer (1988) have discussed possible topologically-distinct spatial relations between two objects in a 1-dimensional space. Egenhofer discussed this in 2 dimensions at the NCGIA workshop on "Languages of Spatial Relations" (see Mark et al., 1989, p. 24, and Figure 8, p. 21). For the expression of relative locations in natural language, however, the figure object (object to be located) is conceptually almost always considered to be a point (Talmy, 1985), or at least to be of insignificant extent. In such a case, when the figure and ground are specified, and when the ground is a "geographic area", five situations can be distinguished geometrically (see Figure 1). These situations might be distinguished in natural languages as well.

In situation "A", the English preposition "in" will often be appropriate ("Caracas is in Venezuela"), although some ground objects require "on" ("his house is on the military base"). In situations "B", "C", and "D", the figure is contiguous with the boundary of the ground object. If the figure is not a point, then we might distinguish three situations, depending on whether the figure is continuous with the inside edge of the boundary (B), straddles the boundary (C), or is against the outside of the boundary (D). Finally, the figure may be disjoint from the ground, yet may still be located with respect to it (Figure 1, E).

When the ground object is a linear feature (such as a river or a road), there are only three distinct situations (see Figure 2).

As an example of potential difficulty in the translation of an apparently-simple situation, suppose that a procedure has been successfully written to determine when to use the English word "on" to describe geographical situations. (We note that the writing of such a procedure would not be a trivial task!) If the procedure has been designed with only English in mind, then translation to German or Spanish could present different difficulties. In German, "on" becomes two distinct concepts, represented by **auf** and **an**, where **auf** represents, for example, the relation between a book and a table, and **an** represents the typical relation between a picture and a wall. In geographic space, German would use **an** for relations where the figure object is adjacent to a water body or transportation medium, situations such as those shown in Figures 1D and 2B. Perhaps the hypothetical English-based procedure would not properly distinguish between these cases. On the other hand, most geographical uses of "on" would become **en** in Spanish, but the Spanish **en** also includes many cases where English would use "in". Furthermore, **sobre** may occasionally be used for "on" situations represented by the German **auf**, but never would be substituted for the German preposition **an**.

Although this example is very simple, it could arise if a GIS user queried a database for dwelling places "on a lake". In German, depending whether **an** or **auf** was used, the system might be expected to return residential parcels adjacent to the lake or, alternatively, houseboats within the lake's boundary. Mismatches of fundamental spatial concepts are expected to be more pronounced for at least some non-Indo-European languages. Much of what is discovered during research on this topic may have a significant impact upon the strategies used to define spatial objects and their relations in object-oriented databases. Some differences in the use of locative prepositions in English and Spanish will be discussed in detail below.

A computer system capable of "understanding" some knowledge domain, if available for more than one natural language, would be capable of translating from one of those natural languages to another. Our approach to the topic of cross-linguistic studies of geographic concepts has as a central theme an attempt to identify language universals and language primitives for geographic space. These would be the essence, the kernel, of an object-oriented central database. Language-understanding or language-production modules then would translate between various natural languages and the central database. The idea of using a neutral, central model in machine translation between natural languages is not an original one. In fact, the Distributed Language Translation Project, at Utrecht, Netherlands, uses exactly such an architecture, with Esperanto (an artificial language invented in the 19th Century in an attempt to reduce cross-linguistic difficulties) as the intermediate language (Schubert, 1986, 1988). From the perspective of GIS and geographical analysis, identification of core of concepts for geographic objects, relations, and information is very high on the research agenda for geographic language.

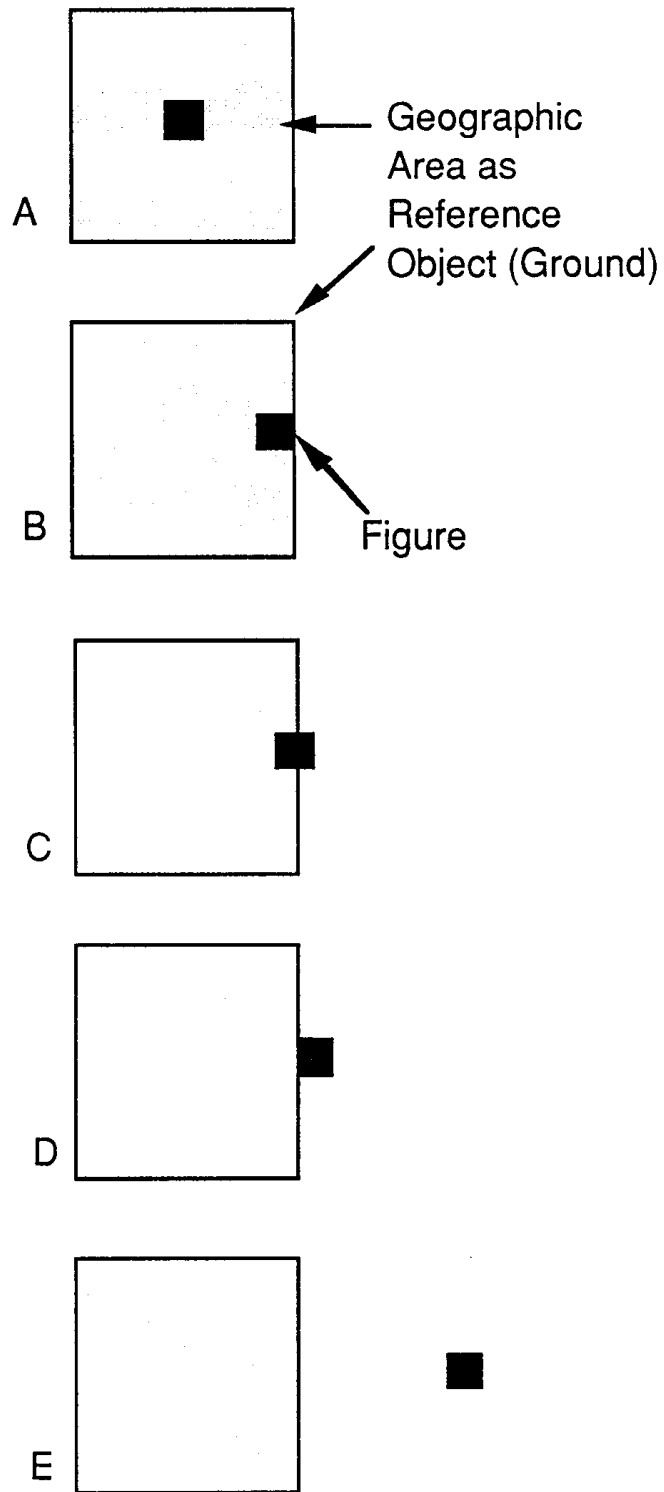


Figure 1: Possible topologically-distinct relations between a point, or relatively small figure object and a reference (ground) object which is an extended geographic area.

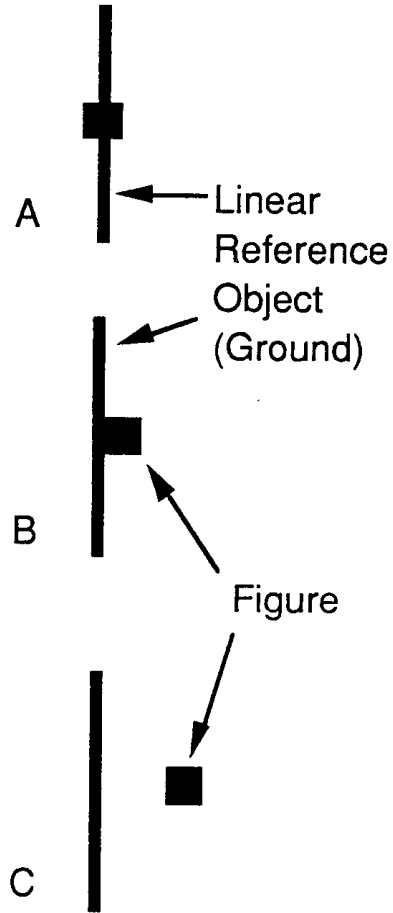


Figure 2: Possible topologically-distinct relations between a point (or at least, small) figure object and a reference object (ground) which is a linear geographic feature.

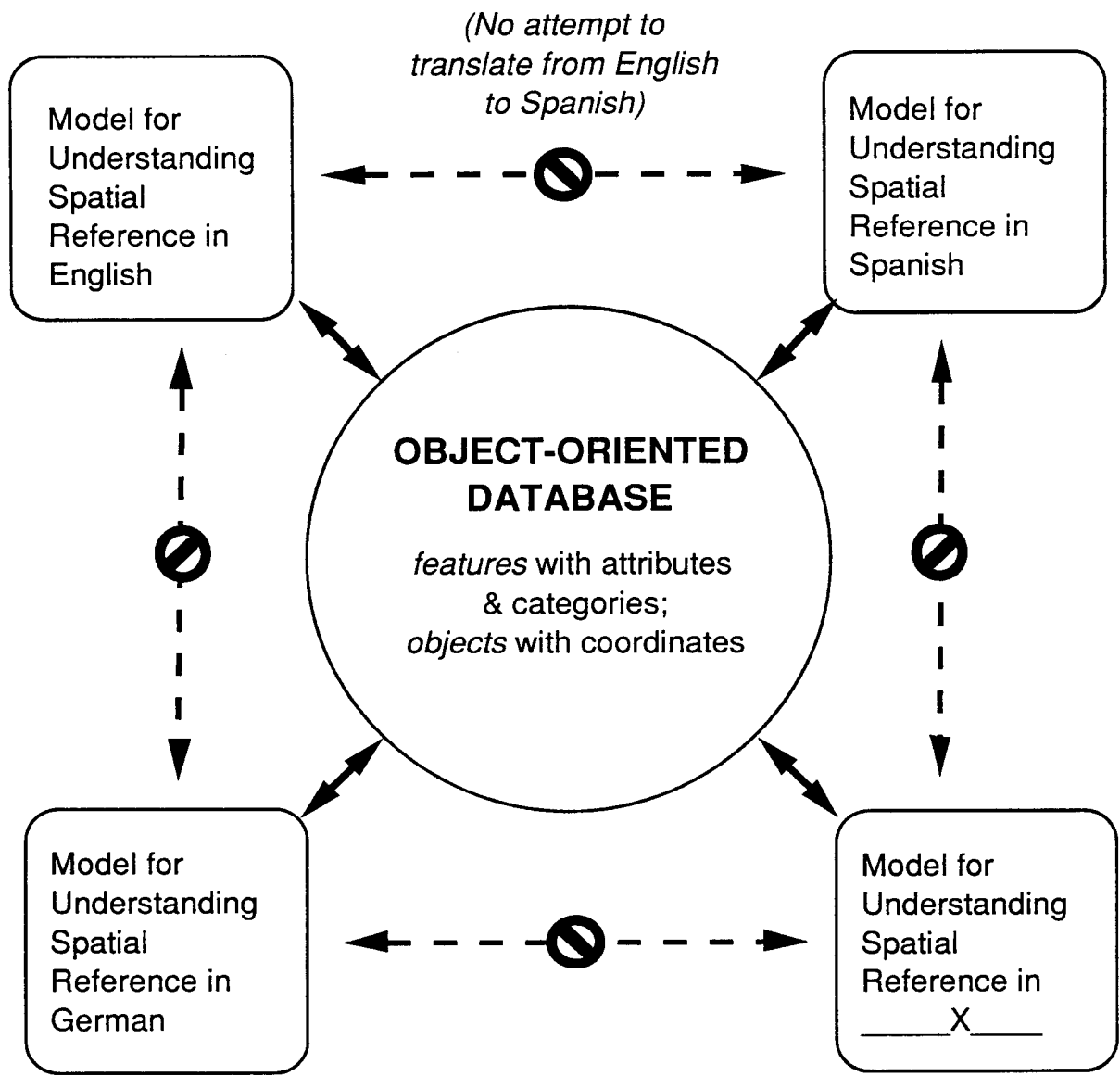


Figure 3: Conceptual model for an approach to machine translation of geographic terms and concepts.

The idea of using a neutral, central model in machine translation between natural languages is not an original one. In fact, the Distributed Language Translation Project, at Utrecht, Netherlands, uses exactly such an architecture, with Esperanto (an artificial language invented in the 19th century in an attempt to reduce cross-linguistic difficulties) as the intermediate language (Schubert, 1986, 1988). From the perspective of GIS and geographical analysis, identification of core concepts for geographic objects, relations, and information is very high on the research agenda for geographic language.

HUMAN-COMPUTER INTERACTION (USER INTERFACES)

Human factors research, part of which is dedicated to human-computer interaction (HCI), has begun to focus attention on cross-linguistic and cross-cultural aspects of software system design. An exemplary product of this attention are foreign language versions of American -designed word processors. Some new products allow for vertical (Asian languages) or right-to-left (Hebrew) text formatting, while others have made allowances for differing hyphenation methods (as is needed for Spanish). Far more research and testing is needed, however, in order to begin to develop systems with salient and intuitive command names in various languages. Part of the problem rests in the inability of certain languages to synthesize complex, multi-word concepts, which makes short, concise menu titles difficult to develop. This is partly true of both Spanish and German.

Although it is obvious that GIS terminology will prove far more complex and linguistically demanding, a cross-linguistic design initiative, similar to that underway for word processors, should be undertaken. Work on computer user interfaces for international technology transfer is already a focus of research attention (see Nielson, in press; especially Sukaviriya and Moran, in press).

Menu Names

To get an impression of the difficulties associated with cross-linguistic human-computer interaction, consider the translation of simple menu items and system commands. These are usually verbs of the English language, but often are abbreviated to conserve menu space. However, many other human languages do not allow such liberty in describing complex relations and actions.

An example problem was discovered during a recent visit to a transportation-planning agency in Barcelona, Spain. Users of the Map Analysis Package (MAP) described the English-language interface as "extremely challenging" to work with, given the "unnatural" syntax and strange vocabulary, but once the commands have been learned the users' performance was quite adequate. One user relied heavily upon a four-inch-thick printout of FORTRAN source code to determine the meanings of many commands. The "SPREAD" function (which determines a zone of all points within a user-specified distance of some geographic object, a function called "buffering" in some other GISs) would be properly translated to the verb **extendar** (to extend). However, this conflicts with another MAP command, "EXTEND". An alternative in Catalan is the verb **escampar**, which in Peninsular Spanish is used to describe a scattering, rather than a continuous zone or region as computed by "SPREAD". (We use "Peninsular" Spanish to refer the version or dialect of the language that commonly is spoken in Spain, as opposed to in the Americas.) It is interesting to note that while the MAP *users' manual* was translated completely into Catalan by the transportation agency six or seven years ago, no attempt was made to translate the *command names* into Catalan, even though the source code and the programming expertise were available.

Another system, written in-house by a city government planning office in Catalufia for cadastral mapping, offers Catalan and Castilian (Peninsular Spanish) language options. In this system, command options are listed, and may be chosen using numeric keys. Little attempt at command abbreviation has been made, and some commands approach 80 characters in length and have a sentence-like structure. Still, not all commands retain their meanings. For example, one of the command options is <<**Digitalizacion nube de puntas**>> ("digitize a cloud of points"). Similar command translation problems exist for German as well, which tends to produce long, compound words that are not easily abbreviated.

SPECIFYING GEOGRAPHICAL LOCATIONS IN SPANISH AND ENGLISH: SOME EXAMPLES

In this section, we present a discussion of a few aspects of the use of prepositions to specify geographical locations in Spanish and English. We recognize that the examples we present are only vaguely related to GIS problems and queries, but believe that the examples give at least a hint of the complexity of the cross-linguistic specification of geographical relations, even between two Indo-European languages. Also, we wish to point out that we by no means claim that this analysis is exhaustive, and that we hope that readers of this material will inform us of any errors or misconceptions in our comments on locative expressions in Spanish and other languages.

In the book "Modern Spanish", the Spanish equivalents of the English preposition "at" are introduced as follows (TMLA, 1973, p. 368):

at	<i>place in which (no motion toward)</i>	en
	<i>point in space, time, or on a scale a place to or toward which</i>	a

This introductory textbook on Spanish for English speakers continues:

"**En** is the equivalent of *at* when the place (or event) is thought of as having dimensions within which something is located (or occurs). In most other instances the equivalent of *at* is **a**." (TMLA, p. 368).

The authors then go on to give a fairly detailed discussion of how motion and scale influence the choice between **en** and **a** for Spanish locative expressions. Apparently, the choice in Spanish is dominated by a static-dynamic component: **a** is required for destinations, intersections, paths, and trajectories, whereas **en** is commonly used for static co-occurrence in space. English, on the other hand, chooses between "at" and "in" mainly based on whether the reference (ground) object is conceptually collapsed to a point ("at") or not ("in"). Of course, this "collapsing to a point" is itself induced by a dynamic situation, and so the uses end up being similar in many cases. This static-dynamic dichotomy permeates the Spanish language. The verbs **ser** and **estar** are prime examples-- the former is used for permanent conditions, the latter for temporary ones.

En versus **sobre**, **dentro**, etc.:

At first examination, it appears that, in Spanish, the preposition **en** is used to specify most geographic and many other spatial relations. However, it appears that in Spanish one must distinguish between two types of locative information: **where** something is located, and **how** it is located. To specify where something is, Spanish almost always uses the preposition **en**, which functions more or less as a "generic" locative. But when a Spanish speaker wishes to express **how** something is located in geographic space, they often will use **sobre**, **dentro**, or some other preposition commonly used for specifying locations of everyday objects.

In some cases, however, both situations fuse into a single usual expression, which is more proper than the generic one. For example, one Spanish speaker felt that **la casa sobre la colina** ("the house on the hill") was more usual in Peninsular Spanish than **la casa en la colina** or **la casa de la colina**; however, another, from a different part of Spain, disagreed, stating that he preferred **de**. It will depend of course of the context or the intention of the speaker, and on the details of the location. **La casa sobre la colina** means, more or less, "the house on top of the hill," but in Spain the house does not need to be at the exact summit of the hill for **sobre** to be appropriate. Of course, if the house is closer to the base of the hill or mountain than it is to the top, **sobre la colina** is not allowed. In such a situation, a speaker of Peninsular Spanish would say **en** (or **de**) **la falda de la colina** ("on the brow of the hill") if the figure object is on the side of the hill, away from the top. Jeff Lansing

(personal electronic mail communication, 6 July 1989) elicited verbal responses from two native Colombian speakers of Spanish, by showing the subjects a photograph. He reported **en la colina** to be the first choice, with **sobre la colina** apparently meaning precisely at the top. Furthermore, Lansing reported that, in the dialect he studied, **en la falda** would only be used by "people who know geography".

Similarly, while **dentro** generally means "in", it is more nearly equivalent to "within" or "inside of", and seems to be used in geographical situations even less often than **sobre**. An example of the geographical use of **dentro** is: **dentro del bosque** (meaning "in the woods"); however, **en el bosque** is perfectly acceptable.

This seems to be one of the main differences between geographical locatives in Spanish and English, more important than the fact that Spanish does not use several different prepositions for different spatial relations or that it uses, in general, just one preposition for geographic space. The lack of compulsion and the distinction between where things are located and how things are located could perhaps be explained by the fact that Spanish tends to express many locational situations as if they were the results of previous actions or events (either actual or virtual). "Where" questions seem to be satisfactorily answered by a generic statement, which does not require explicit mention of the specific spatial relation, provided that the question is not asking for those details.

This difference between Spanish and English has potential practical significance for GIS, since the two types of locative questions recognized in Spanish are not often distinguished in English. Thus menus and systems written by English-speakers will be unlikely to provide the user with a choice between "where" and "how" things are located. They also may provide inadequate or incomplete answers to locative questions thought out and composed in Spanish, or in other languages with similar spatial representation systems.

A Use Type for the English *on*

As a further example of cross-linguistic complexity, we will discuss just one of Herskovits' (1987, p. 148) "use types" for the English preposition *on*: "Physical object contiguous with edge of Geographic area."

Some of Herskovits'(1987, p. 148) examples for this use type in English are:

a shop on the main square
a garden on the lake
a city on the ocean
the store on Polk street
a room on the patio
a house on the park
the gas station on the freeway

Herskovits notes that for this use type in English, the reference object, a "geographic area", must be among the following three categories: "open spaces, bodies of water, and all manner of pathways" (Herskovits, 1987, p. 148). (In English, for any other type of extended geographic reference object, the edge must be explicitly mentioned.) As we will discuss below, the situations expressed in the above examples are not expressed in a uniform way in Spanish.

When a location "on" a water body is expressed in Spanish, a noun such as **orilla** ("shore" or "bank") is required as part of the expression. Phrases such as **a la orilla de** or **en la orilla de** apply to all types of water bodies, including the ocean, seas, rivers (including even minor streams), and lakes. One informant claimed that, in Peninsular Spanish, **orilla** cannot be applied to situations adjacent to the edges of other extended reference

objects, such as roads, streets, or plazas. However, another informant from Spain reports that the expressions **a la orilla del camino** and **a la orilla de la carretera** are indeed very common, not only for example in literature, but in daily speech as well. Note that the preposition used is **a** and not **en**.

Jeff Lansing (personal electronic mail communication, 12 July 1989) gave some examples of this from the Spanish of Colombia:

Mi casa es en la plaza "My house is on the edge of the plaza."

Mi casa es en la mitad de la plaza "My house is out in the (middle of the) plaza."

Mi casa es en la orilla de la laguna "My house is on the edge of the lake."

Mi casa es en la laguna "My house is out in the (middle of the) lake" (plausible if it was built on some rocks sticking up in the middle of the lake).

The two situations, the plaza and the lake, seem similar, and equally unambiguous. Typically, houses are not "within the spatial polygons called" either lakes or plazas, but would be next to them. Typicality or conventionality would account for the plaza example since the Spanish preposition *en* seems almost to mean, "is located with reference to". The typical relation of a house to a plaza (assuming that the plaza is the salient reference object) is adjacency. One might expect language to "mark" the atypical situation of a house out on the plaza linguistically, just as the example <<**Mi casa es en la mitad de la plaza**>>, above. But, if we apply that reasoning to the lake, the same "should" hold-- a house within the boundaries of a lake certainly would be unusual, and so we might expect "a house <<on>> the lake to be adjacent to it.

In Peninsular Spanish, **orilla** is restricted to water bodies, and thus Spaniards definitely cannot say **a la orilla de la colina** ("on the 'edge' of the hill"). Jeff Lansing (personal electronic mail communication, 6 July 1989), however, states that **a la orilla de la colina** is quite acceptable in an American Spanish dialect spoken in Bogota, Colombia. (This is perhaps a general difference in usage between American and Peninsular Spanish, but we need more information before we can state this for certain.)

Apparently, there is not a specific Spanish word for river edges which is different than that for seas; however, one often uses the expression **en la costa** ("on the coast"), which is only used with respect to seas, and which implies the mention of the sea itself. Thus, it is roughly equivalent to say **a la orilla del mar** or **en la costa**, although it should be noted that **la costa** is more a fringe or strip concept, a region of land near the sea, whereas **orilla** refers more strictly to the edge, the boundary between land and sea. According to Spanish dictionaries, **ribera** applies indifferently to rivers and seas or lakes; however, in practice it is used more often for rivers than oceans because of the "literary" connotations. Moreover, it appears as a place name for places located near to rivers, especially as a generic place name, **La Ribera** (in Navarra). At least in eastern Spain, **ribera** is more "literary", and not heard often in everyday speech. But at least in Galician, the equivalent word **ribeira** is, for fishermen, as common as "interface" is for a computer scientist. Spanish also has another more general word, **borde**, that can be used for the edges of rivers, lakes and oceans.

Other Romance Languages

We have studied some other Romance languages in less detail, but present a few of the facts and examples we have collected so far.

Like Spanish, Catalan has genetically related edge-words, **riba** and **ribera**, which apply also indifferently to both rivers and oceans. More complicated is the case of the Catalan **vora**, which means "edge" and is applied in both geographical and other contexts. At the same time it is also the Catalan word for **orilla**. Finally, and to complicate matters further, the Catalan **vora** is also the equivalent of "near" or **cerca** (Spanish). Catalan also has another word for this concept (**prop** or **a prop**) but both **vora** and **prop** are used interchangeably in Catalan.

In French, the edge often is mentioned explicitly when the reference object is a water body. **Rive** ("bank") is used commonly for the edge of a river or stream, and also for lakes or the ocean. However, the more general **bord** ("edge") seems to be used more often for the ocean. **Oree**, another word meaning "edge", genetically related to **orilla**, and derived from the Latin **ora**, might be preferred over **borde** or **rive** for the edge of a wood. In European French, the edge noun, while common, apparently is not obligatory, whereas in Quebec it unusual to omit it.

In Rumanian, one commonly uses an edge noun in the cases for a generic reference water body (**Casa de pe malul lacului**, "The house on the lake"; **mal** is a Rumanian word representing "shore"), although it is not strictly required (Lior Moscovici, Internet electronic news posting, sci.lang article 4904, 14 July 1989). However, Rumanian more often omits explicit reference to the edge of the water if you refer to a specific body of water (**Casa de pe lacul X**, "The house on Lake V). Latin apparently did not require edge nouns.

All of these examples indicate that Herskovits' use type "physical object contiguous with edge of Geographic area" is not a fundamental (primitive) cross-linguistic type of spatial relation, since in Spanish some cases (water bodies) require a noun phrase that explicitly mentions the edge or boundary, whereas others (for example, roads) do not. This might have been anticipated from an examination of English alone, since in English one can only use a geographic area without an edge noun as a ground or reference object if the object has certain specific attributes.

Prepositions for Geographic Relations

The choice of the preposition *before* **orilla** (that is, the choice between **a la orilla de** and **en la orilla de** does not seem to depend on the nature of the spatial relation or of the reference object. (See discussion of **en** versus **a**, above.) The use of **a** or **en** is, in many cases, ruled by the particular verb that precedes the preposition, and whether it implies motion or a static situation. Some verbs require **a** (**estar**, for instance), while others take **en** (for example, **situar**). Similarly, in Catalan, one must choose between **a** and **en**, but mostly on a phonetic basis, although it seems to prefer **a** in locatives which include specific place names.

Another example shows the role of pronunciation in the choice of preposition in Spanish. When a house is located very near to the base of a hill, a Spanish speaker would say **la casa al pie de la colina**, which also applies to buildings or any other body which has some height compared with the object located. This complements the expression **la casa en** [never **a**] **la falda de la colina**. In some of those expressions, including words such as **pie**, **falda**, **cima**, or **orilla**, apart from the dynamic/static rule to decide between **a** and **en**, there is also a pragmatic question: if the article that follows is male (**el**), Spanish can contract **a** but not **en**, and conversely the contraction never can be done if the article is female (**la**). The rules favor the contraction situation, and thus it is **al pie**, yet **en la falda**.

DISCUSSION AND PROSPECTS

The foregoing was a long excursion into cognitive science and linguistics. We now conclude by discussing research issues more directly related to GIS, to computer interfaces and data structures, and to cross-cultural technology transfer.

Present GIS design allows for the use of technology only by a minority consisting of highly-trained personnel. This is true even in English-speaking countries, but is amplified many-fold in countries where other languages are spoken. Based on some anecdotal evidence, it seems that people who use "English" software (that is the *current* users) accept this situation as *status quo*. Also, for some users, there probably exists an element of elitism associated with *being able to understand English* well enough to operate computer software. Software written and supported in Spanish would allow "anyone" to be a computer user. If knowledge is power, control of information technology also is power. Furthermore, many people find documentation and software to be their best (perhaps only) chance to learn (or practice) their English.

We contend that the application of creative and innovative science, engineering, art, and application in the geographical sciences is hampered by the present need for users to devote a considerable majority of their time to building, learning, optimizing, and maintaining systems and databases. Again, this effect is exaggerated when translation of manuals, documents, and commands is also needed.

It is not too soon to begin designing systems to take advantage of real-time spoken natural-language understanding. Systems are already able to understand such speech within limited knowledge domains and for limited vocabularies. The technology probably will be available before we can implement and test language - understanding procedures for GIS and spatial data handling.

Lastly, a goal of GIS design should be to liberate GIS from the laboratory, and to allow its introduction into mainstream life-- in public libraries, hotels, airports, taxis, and even in private homes. After all, similar technology, such as word processors and spreadsheets, already can be found in many homes in the industrialized nations.

REFERENCES

- Herskovits, A., 1985. Semantics and pragmatics of locative expressions. *Cognitive Science* 9, 341-378.
- Herskovits, A., 1987. *Spatial prepositions in English*. Cambridge, Mass.: Cambridge University Press.
- Johnson, M., 1988. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason..* Chicago, IL: University of Chicago Press.
- Lakoff, G., 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago, IL: University of Chicago Press.
- Ma, P., 1987. An algorithm to generate verbal instructions for vehicle navigation using a geographic database." *The East Lakes Geographer* 22: 44-60.
- Mark, D. M., 1985. Finding simple routes: 'Ease of description' as an objective function in automated route selection. *Proceedings, Second Conference on Artificial Intelligence Applications (IEEE)*, Miami Beach, December 11-13, 1985, pp 577-581.
- Mark, D. M., 1988. *Cognitive and Linguistic Aspects of Geographic Space: Report on a Workshop*. Santa Barbara, CA: National Center for Geographic Information and Analysis.
- Mark, D. M., 1989. *Languages of Spatial Relations: Researchable Questions and NCGIA Research Agenda*. Santa Barbara, CA: National Center for Geographic Information and Analysis, Report 89-2A.
- Mark, D. M. and Frank, A. U., 1989. Concepts of space and spatial language. *Proceedings, Ninth International Symposium on Computer-Assisted Cartography (Auto-Carto 9)*, Baltimore, Maryland, pp. 538-556.
- Mark, D. M., Frank, A. U., Egenhofer, M. J., Freundschuh, S. M., McGranaghan, M., and White, R. M., 1989. *Languages of Spatial Relations: Initiative Two Specialist Meeting Report*. Santa Barbara, CA: National Center for Geographic Information and Analysis, Report 89-2.
- Mark, D. M., Svorou, S., and Zubin, D., 1987. Spatial terms and spatial concepts: Geographic, cognitive, and linguistic perspectives. *Proceedings, International Symposium on Geographic Information Systems: The Research Agenda*, November, 1987, Crystal City, Virginia, vol. 2, pp. 161-112.
- McGranaghan, M., 1989. Context-free recursive-descent parsing of location-description text. *Proceedings, Ninth International Symposium on Computer-Assisted Cartography (Auto-Carto 9)*, Baltimore, Maryland, pp. 580-587.
- McGranaghan, M., Mark, D. M., and Gould, M. D., 1987. Automated provision of navigation assistance to drivers. *The American Cartographer*, 14: 121-138.
- Nielson, Jakob, editor. *Designing Interfaces for International Use*. Elsevier Science Publishers, in press.
- Pullar, David V., and Egenhofer, Max J., 1988. Toward formal definitions of topological relations among spatial objects. *Proceedings, Third International Symposium on Spatial Data Handling*, Sydney, Australia, August 17-19, 1988, 225-241.

- Schubert, Klaus, 1986. Linguistic and extra-linguistic knowledge. A catalogue of language-related rules and their computational application in machine translation. *Computers and Translation* 1, pp. 125-152
- Schubert, Klaus, 1988. Ausdruckskraft und Regelmässigkeit: Was Esperanto fuer automatische Uebersetzung geeignet macht. *Language Problems and Language Planning* 12, pp. 130-147
- Streeter, L.A., Vitello, D., and Wonsiewicz, S.A., 1985. How to tell people where to go: Comparing navigational aids. *International Journal of Man/Machine Studies* 22, 549-462.
- Sukaviriya, Pinyawadee, and Moran, Lucy, in press. User interfaces for Asian countries. In Nielson, Jakob, editor. *Designing Interfaces for International Use*. Elsevier Science Publishers, in press.
- Talmy, L., 1983: How language structures space. In H. Pick and L. Acredolo (eds) *Spatial Orientation: Theory, Research and Application*. Plenum Press.
- TMLA (The Modern Language Association), 1973. *Modern Spanish* (Third Edition, 1973), New York: Harcourt Brace Jovanovich, Inc.