A Methodology for Analyzing Terminological and Conceptual Differences in Language Use across Communities

Mildred L. G. Shaw
Brian R. Gaines

Abstract

The study of discourse within specialized communities with a common disciplinary, or goal–directed, focus requires methodologies for the empirical derivation and comparative analysis of both the conceptual and terminological structures used by participants. This paper reports on research studies over the past sixteen years in which we have developed a theoretical framework, quantitative empirical methodologies, and computer–based analytical tools, for modeling terminological and conceptual relations in language use across communities. In particular, recent developments are illustrated of on–line, computer–based tools for the interactive elicitation of concepts and terminology using graphic interfaces on personal computers. Some new extensions allowing the use of these tools over networks to support discourse and knowledge processes in dispersed communities are described. The presentation emphasizes the close links between the theoretical foundations, linguistic and knowledge representation concepts, empirical methodology, its implementation in computer–based tools, and practical applications. It illustrates the powerful tools now available for the study of quantitative linguistic phenomena in specialized communities.

Introduction

This paper presents an overview and examples of a line of long–term research studies in which we have attempted to develop theoretical foundations, quantitative methodologies, and computer–based tools to elicit, model and compare the conceptual structures of individuals and the terms used in discourse based on them. There are deep theoretical and empirical issues associated with the nature and status of conceptual structures, the interplay between them and language, and the individual and mutual processes involved in discourse. Consequently, this paper presents the theory and methodology in a minimalist fashion, as an experimental and analytical approach with minimal preconceptions. We see the methodologies and tools having a wide range of applications that, to some extent, are independent of particular theoretical positions. However, no measuring instrument is free of artefacts, and no application of one does not perturb the situation measured; hence, we shall attempt to delineate clearly the essential preconceptions of the methodologies and tools.

R. Köhler and B.B. Rieger (eds.), Contributions to Quantitative Linguistics, 91–138.
© 1993 *Kluwer Academic Publishers. Printed in the Netherlands.*

The roots of these studies are in our fascination with the dynamic interplay between the anarchic individualism of the personal conceptual world on one hand, and apparently well–defined consensualism of social conceptual worlds on the other, particularly those of science as *objective knowledge* independent of individuals (POPPER 1968). Whether one adopts a *psychological perspective* and wonders how an individual comes to standardize at least part of his or her conceptual structures and terminology to engage in meaningful discourse with others, or a *knowledge perspective* and wonders how consensual conceptual structures arise out of the pooling of individual experience, or a *linguistic perspective* and wonders about the role, origins and status of language in personal, social and scientific life, there is a profound sense of there being an overall *system* that transcends any of these perspectives. There is then, naturally, a desire to investigate, probe, and measure that system and analyze its processes.

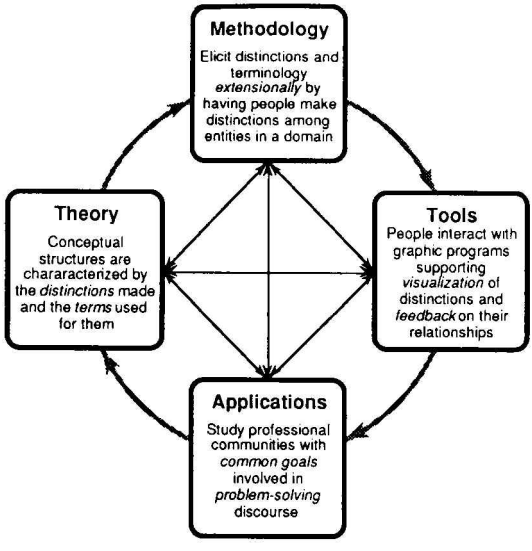


Fig.1 Theory, Methodology, Tools and Applications

We gave a detailed account of the research studies in a linguistic framework a decade ago in BURGHARD RIEGER's (1981) collection on Empirical Semantics (SHAW; GAINES 1981). This paper updates all aspects of that paper, notably in terms of: an extended and refined theory, still embedded in cognitive and social psychology but also firmly rooted in logic and knowledge representation; multi–faceted methodologies accessing knowledge structures at many levels through many paths; improved interactive tools reflecting the improvements in speed and graphics of computer workstations; and a wide range of application studies in collaboration with others worldwide, particularly in the development of knowledge–based systems.

Figure 1 illustrates the main dynamics of the research, that there is a cycle from fundamental theory, through its application and testing as a measurement and modeling methodology, supported by interactive computer tools providing adaptive, on–line experimentation, to a diversity of applications in science and industry. Across the cycle there is continuous feedback from one aspect of the research to others, for example, that methodology is tested in applications, that applications in themselves often lead to extensions of theory, and so on. The relation between the tools and the other domains is particularly interesting. The computer has provided modern science with a vehicle with which to operationalize theory, to make the consequences of theory manifest through exact simulation and principled application. This has become a very important dimension of our research.

Section 2, following, gives an overview of theory; Section 3, methodologies for the elicitation and comparison of conceptual systems; Section 4, tools supporting these methodologies; Section 5, a detailed methodology to derive conceptual relations within a group; and Section 6, an overview of recent developments.

UNDERLYING COGNITIVE AND SOCIAL THEORY

There is no widely agreed model of the underlying dynamics of human socio–cognitive processes — cognitive and organizational psychology research is an area of explosive growth (GAINES 1991a), particularly as our understanding of the modern era is enhanced by post–modern perspectives (HELLER 1990). This section outlines the theoretical perspectives underlying the studies reported in this paper.

There are three basic premises underlying our work:

- that knowledge processes in society can be understood as the modeling procedures of a distributed anticipatory system;
- that these modeling procedures can themselves be modeled in terms of the distinctions made;
- that the cognitive processes of individuals and organizations can be understood within this framework.

The Human Species as a Cognitive System

The distributed anticipatory system perspective takes the human species as a cognitive unit, seeing anticipation of future events as a major species trait. The species' cognitive processes are distributed across its subunits, from social structures to roles in individuals. Language is a mechanism for coordination among the distributed units, and media are a way of achieving coordination over space and time (GAINES 1989a, 1991a).

Simmel made this inter–relation of wholes and parts the center piece of his sociology:

> Society strives to be a whole, an organic unit of which the individuals must be mere members. Society asks of the individual that he employ all his strength in the service of the special functions which he has to exercise as a member of it; that he so modify himself as to become the most suitable vehicle for this function...man has the capacity to decompose himself into parts and to feel any one of these as his proper self. (SIMMEL 1950:58-59)

Simmel's insight that the group member is always a fragment of a person, a role created precisely to enable the person to enter the group has been developed extensively by Wolff with his notions of *surrender* and *catch*:

> From the standpoint of the world of everyday life, the mathematician, as we often put it, lives in the 'world of mathematics', dealing with 'nonreal' elements, notably numbers, whose relation to 'real' things, to 'reality', is not part of his concern. Analogously for the logician. What makes our subject–object approach to this attitude misleading is the fact that the subject, the student of mathematics or logic — his or her individuality, including motives and attitudes — is irrelevant for our understanding; the only thing that counts is the pursuit, with its results and questions. (WOLFF 1976:162-163)

He makes the key point that not only does the real world, the object, disappear to be replaced by the world of mathematics, but also that in entering into this world the person doing mathematics, the subject, also disappears to be replaced by a new entity, the mathematician.

Simmel and Wolff were not alone in these insights and there are two further models which develop and complement theirs. In terms of Simmel's parts and Wolff's subject PASK's (1975) concept of "P–individuals" as coherent psychological processes capable of engaging in conversations is a useful representation of the results of Simmel's modifications and Wolff's surrender, particularly when we note that several P–individuals may execute within a single processor. Thus 'the mathematician' may engage in conversation with 'the physicist' or 'the statesman' all of whom happen to use the same brain for their processing.

In terms of Wolff's subject–object distinction, his "object" corresponds to notions of *knowledge*, and is captured by POPPER's (1968) concept of a "third world" of "statements in themselves". The third world of knowledge is a useful representation of that which we "catch", particularly when we note its distinct ontological status:

> I regard the third world as being essentially the product of the human mind. It is we who create third–world objects. That these objects have their own inherent or autonomous laws which create unintended and unforeseeable consequences is only one instance (although a very interesting one) of a more general rule, the rule that all our actions have such consequences. (POPPER 1974:148)

Pask goes beyond Simmel in conceiving that a P–Individual is not just what we conventionally term a 'role' within a person but may itself be composed of a number of roles coming together to form a unity that we conventionally term a group or organization:

> a P–Individual...has many of the properties ascribed by anthropologists to a role, in society or industry, for

example. A P-Individual is also a procedure and, as such, is run or executed in some M-Individual, qua processor. However, it is quite exceptional to discover the (usually assumed) one to one correspondence between M-Individuals and P-Individuals. (PASK 1975:302)

SHAW (1985) has developed Pask's notions within the framework of personal construct psychology to show how Kelly's cognitive psychology may be used to account for the psychological processes not only of individual people but also for that of functional groups such as a nuclear family or a product executive. Figure 2 illustrates these ideas. Anne is shown as having three roles, wife, mathematician and technical vice president. John is also shown as having three roles, husband, golfer and sales vice president. The wife role in Anne together with the husband role in John together constitute a new P-Individual or cognitive entity, a nuclear family. This has behavior, language, legal rights, and a model of the world, that are distinct from those of Anne and John in their other roles, and goes beyond those of the participating roles in Anne or John alone. Similarly, the technical VP role in Anne, the sales VP role in John, the finance VP role in Mark and the production VP role in Sue constitute another P-Individual or cognitive entity with again behavior, language, rights and models that are distinct from its participants and other cognitive entities in which they participate in other roles.

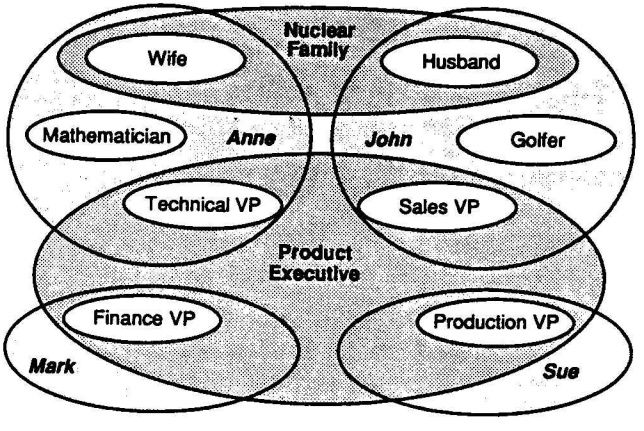


Fig.2 Roles and groups as cognitive systems of psychological individuals

Cognition as a System of Distinctions

The theoretical foundations for developing, eliciting, analyzing and comparing the conceptual frameworks of the cognitive entities in Figure 2 is derived from the notion that concepts correspond to terms applied to *distinctions* that we make in modeling a world:

a universe comes into being when a space is severed or taken apart...By tracing the way we represent such a severance, we can begin to reconstruct, with an accuracy and coverage that appear almost uncanny, the

basic forms underlying linguistic, mathematical, physical and biological science, and can begin to see how the familiar laws of our own experience follow inexorably from the original act of severance. (BROWN 1969)

We have shown that a wide range of psychological and philosophical frameworks are characterized by the systems of distinctions underlying them, and that there are a number of basic distinctions common to such systems (GAINES; SHAW 1984). The frameworks include that of KELLY's (1955) personal construct psychology, and his notion of "personal constructs" as filters through which a person perceives events is captured abstractly through the notion of a distinction. Kelly makes the same observation as Brown but couched in psychological terms:

Man looks at his world through transparent templets which he creates and then attempts to fit over the realities of which the world is composed. (KELLY 1955)

He continually emphasizes the epistemological status of these constructs in predicting and controlling the world and their ontological status as personal conjectures rather than reality–derived absolutes:

Constructs are used for predictions of things to come, and the world keeps on rolling on and revealing these predictions to be either correct or misleading. This fact provides the basis for the revision of constructs and, eventually, of whole construct systems. (KELLY 1955)

Formally, a *distinction* is taken to be a primitive notion, but its connotations of cutting out something for some purpose, or because it satisfies a criterion, are informally intended. The connotations that there may be an agent making the distinction, and that something is distinguished are also informally useful. However, it should be noted that *closed–world* assumptions are specifically excluded — it is not assumed that what is *not* distinguished is meaningful or has the same categorial status as that which is distinguished. Informally, only that which is distinguished is known, new distinctions can be made indefinitely, and the world is assumed open to novelty. Formally, this affects the way relations between distinctions are defined — extensional forms, such as "for all", are avoided because it is not reasonable to assume that statements can be made about an indefinite extension whose members are not yet identified and may never be known.

Distinction are the only primitives. One might ask "what is distinguished", but the answer will be "distinctions." This corresponds to deriving the two phenomena in Kelly's *construction corollary*, "a person anticipates events by construing their replications," from a single primitive. Kelly describes the process as

we look at the undifferentiated stream of circumstance flowing past us, and we try to find something about it that repeats itself. Once we have abstracted that property, we have a basis for slicing off chunks of time and reality and holding them up for inspection one at a time...this is done by construing. (KELLY 1955)

That is, we distinguish events in the undifferentiated stream of circumstance, and then we further distinguish among the distinguished events by construing.

A Calculus of Distinctions

In developing a formal calculus of distinctions, we have attempted to specify a minimal set of operators that are adequate both to capture Kelly's cognitive psychology and to provide foundations for KL–ONE–like, term subsumption knowledge representation schemas (GAINES; SHAW 1990). The primary foundations of personal construct psychology are the *organization corollary*, "each person characteristically evolves, for his convenience of anticipating events, a construction system embracing ordinal relationships between constructs," and the *dichotomy corollary*, "a person's construction system is composed of a finite number of dichotomous constructs." In term subsumption knowledge representation systems, these correspond to the notions of one concept subsuming another, and one concept being disjoint to another, respectively.

Formally, one distinction will be said to *subsume* another if it can always be applied whenever the other can. This notion of subsumption turns out to be adequate to capture Kelly's use of the term that one construct subsumes another, and also the use of the same term in knowledge representation, that one concept subsumes another. It can be represented formally as:

$$\text{“}b \text{ subsumes } a\text{”} \qquad a \rightarrow b \;\Leftrightarrow \vdash xa \Rightarrow \vdash xb \tag{1}$$

That is, b subsumes a, if and only if whenever one asserts xa one also asserts xb. Note that the form $\forall x$ is avoided for the reasons given above — that the notion of all the distinctions to which a and b may be applied is not well–defined. However, if this extension is well–defined then the intensional formulation of (1) implies the extensional formulation in terms of $\forall x$. MARCUS's (1962) non–standard, *substitutional* reading of the universal quantifier would be appropriate to the definition above, but the notation is avoided because the extensional reading is prevalent and significant in its own right.

The definition in (1) is best read intensionally in terms of a *commitment* to the way in which distinctions will be made, such that if a is made then there is a commitment to b being made also. Subsumption corresponds to increasing generality since the subsuming distinction can be applied to at least as many things as that subsumed. It is an asymmetric, transitive relation, a partial preorder, over distinctions, that supports the ordinal relations of Kelly's organization corollary.

There is a second relation definable in similar terms which is equally important, that one distinction is disjoint with another in that one can never be applied whenever the other can. This definition of a disjoint relation turns out to be adequate to capture Kelly's notion of a dichotomy, and also the notion of disjoint concepts in knowledge representation. It can be represented formally as:

$$\text{“}a \text{ disjoint } b\text{”} \qquad a\text{—}b \;\Leftrightarrow \vdash xa \Rightarrow \neg \vdash xb \tag{2}$$

That is, a is disjoint with b, if and only if whenever one asserts xa one does not assert xb. The definition is again best read intensionally in terms of a commitment to the way in which distinctions will be made, such that if a is made then there is a

commitment to b not being made. Disjoint is a symmetric, intransitive relation over distinctions, and supports Kelly's dichotomy corollary.

It is interesting to note that definition (2) is an asymmetric definition of what is clearly a symmetric relation. Logically, this is is possible because the reverse implication can be derived from (2), that is, if one asserts xb one cannot assert xa because that would imply $\neg \vdash xb$. This derivation of symmetry from asymmetry may be logically simple, but it is not semantically trivial. In terms of knowledge representation it corresponds to the essential sequence of definitions: if we define a first we cannot define it to be disjoint with b because b is not yet defined. Psychologically, this asymmetry appears to be related to the empirical asymmetries ADAMS–WEBBER (1979) has observed in the use of the, apparently symmetric, poles of a construct.

The $\rightarrow$ and — relations are complementary in establishing four possible binary relations between distinctions, that $a \rightarrow b$, $b \rightarrow a$, a—b, or none of these. The two subsumption relations can hold together giving an equivalence relation on distinctions. The disjoint relation is incompatible with the subsumption relations, and is *inherited* through subsumption, that is:

$$a\text{—}b \text{ and } c \rightarrow a \Rightarrow c\text{—}b \tag{3}$$

These notions are sufficient for the definition of concepts and constructs.

A *concept* is defined to be that mental entity imputed to a distinction making agent as enabling it to make a particular distinction. Note that concepts are separated both from the distinctions they support and the entities they distinguish. This definition corresponds to ANGLIN's (1977:3), "a concept is all of the knowledge possessed by an individual about a category of objects or events" — "Concepts mediate categorization but concepts are not the resultant categories." The notion of an agent and its mentality are deliberately introduced at this stage because they are part of the connotations of a concept. However, consistent, with Kelly's position, concepts are defined to be *imputed* to an agent rather than as a mental structure. They are themselves distinctions made by an observer — possibly, a reflective observer. This is equivalent to ZADEH's (1964) introduction of the notion of the *state* of a system as an imputed parametrization accounting for a system's behavior. Concepts are state variables we impute to a knowledgeable agent.

A *construct* is defined to be a triple consisting of a pair of dichotomous concepts mutually subsumed by a third concept defining their scope of application, what KELLY (1955) terms their *range of convenience*. It is, perhaps, a psychological phenomenon that we tend to conceptualize the world in terms of restricted sorts that are then dichotomized, a phenomenon identified in antiquity (LLOYD 1966) and common across many cultures (MAYBURY–LEWIS; ALMAGOR 1989). However, it is also interesting to note that this restriction of conceptual definitions to apply only to existing sub–sets, giving rise to a well-defined relative negation, is also the key technique in Zermelo–Fraenkel set theory (HATCHER 1982) in avoiding the paradoxes inherent in unrestricted comprehension. Constructs are in some sense, the basic units of psychological conceptualization, and the following section analyzes them in detail.

Figure 3 extends Figure 2 to show the conceptual systems imputed to the roles within people constituting the product executive, and applied by them to make distinctions within a world of products and associated items. We have used POPPER's (1968) *three worlds* notions to label the sub–systems within Figure 3, emphasizing through the "world 3" label that the imputed conceptual systems are modeled within an overt, shared world of discourse, subject to encoding in media. We have termed "world 1" one of *reified systems*, rather than of "physical systems" since it can be any subsystem that is imputed to have some special status of inter–subjective availability, that is, to be "real" in some sense. From the perspective of discourse processes, the elements of a particular world 1 form the *domain* whose construal within the *domain of discourse* that is abstracted in a particular world 3 models the processes of discourse between the cognitive entities forming a community mutually concerned with that domain in a particular world 2.

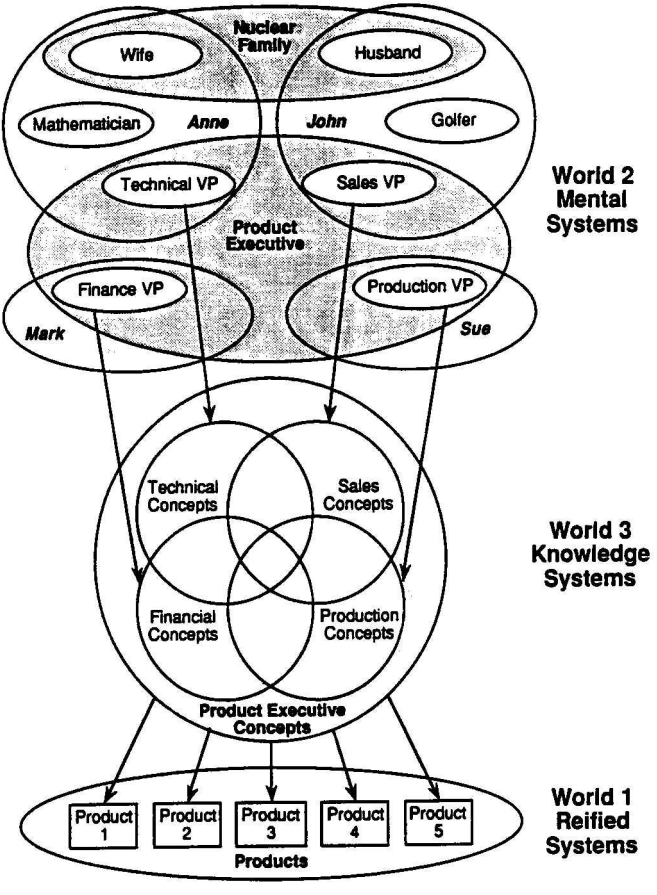


Fig.3 Concept structures as systems of distinctions imputed to cognitive systems to mediate their interaction with a world

The Representation of Distinctions

In developing methodologies and tools, and in presenting knowledge structures, it is useful to develop a graphic representation of the logical relations defined in (1) and (2). Visual presentation of knowledge structures has been an attractive feature of *semantic networks* (QUILLIAN 1968) since their inception. In empirical studies, in particular, the presentation of knowledge structures to those from whom they have been elicited is important to some aspects of their validation. There are many techniques for eliciting knowledge structures but they ultimately result in an operational knowledge base with formal semantics. However, the expression of this knowledge base in the formal language used by the system is usually not very comprehensible to non–programmers (NOSEK; ROTH 1990). A *visual language* that is both comprehensible and formal offers attractive possibilities not only for comprehension but also for editing, and for parts of the elicitation process itself (GAINES 1991b).

We extend the one–dimensional logical language of definitions (1) and (2) by allowing the arrow representing the directed relation of subsumption, and the line to representing the non–directed relation of disjointness, to be at any angle, not just horizontal. Thus, in Figure 4 concept a subsumes concepts b and c, and concepts b and c are disjoint.

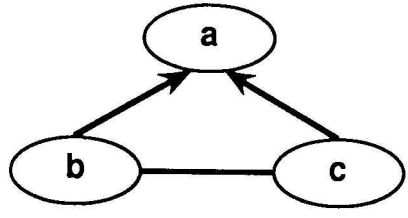


Fig.4 The triple of distinctions generating a construct

Figure 4 also illustrates what for Kelly is the psychological atomic unit, a "personal construct." A *construct* is defined formally to be a triple of two disjoint distinctions mutually subsumed by a third as shown by the structure in Figure 4, and psychologically as the triple of concepts assumed to underlie the distinctions.

The triple captures the notions of *similarity* through the shared subsuming concept, a, and *contrast* through the disjoint subsumed concepts, b and c. The pair of disjoint concepts, b and c, captures Kelly's *dichotomy corollary*, that a person's construction system is composed of a finite number of dichotomous constructs. The concepts b and c characterize what Kelly terms the *poles* of the construct b—c. The subsuming concept captures Kelly's *range corollary*, that a construct is convenient for the anticipation of a finite range of events only. Such a concept, subsuming a dichotomy, characterizes what Kelly terms the *range of convenience* of the construct. The importance of such triples derives from the restricted closed world semantics generated, that the world carved out

by concept *a* is divided by the disjoint concepts *b* and *c* into two parts that are now at the same categorial level.

The subsumes and disjoint relations, → and —, can between them characterize a wide variety of concept and construct structures. The left hand structure in Figure 5 shows that a given range of convenience can be construed through a number of unrelated constructs. The pairs *b*—*c* and *d*—*e* are two different dimensions of the domain characterized by *a*. If *a* characterizes a domain of people then *b*—*c* might be *young*—*old* and *d*—*e* might be *short*—*tall*.

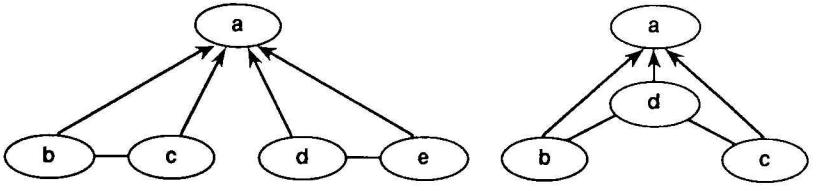


Fig.5 Alternative constructs and alternative poles

The right hand structure in Figure 5 shows that a pole of a construct may be paired with alternative poles that are not mutually disjoint. The pairs *b*—*d* and *c*—*d* are two different dimensions of the domain characterized by *a* that have some commonality through the shared pole, *d*. If *a* characterizes a domain of people then *b*—*d* might be *young*—*old* and *c*—*d* might be *fun*—*old*. It was because different contrasting concepts were possible for the same concept that Kelly saw the concept pairs that form constructs as being a more stable characterization of psychological structure — that the way in which we react to something falling under a concept is dependent on what we see as the contrasting concept.

Linguistics of Distinctions

From a logical and computational point of view the terms which are applied to denote distinctions are just lexical designators which themselves need to be lexically distinguishable, but are otherwise themselves irrelevant in not carrying structural knowledge information. Clearly, from a linguistic perspective this is just not so — the logical and visual languages have made overt in an alternative form structural information which is usually carried in natural language through pragmatic, lexical, syntactic and semantic indicators.

Many linguists and cognitive psychologists have analyzed the relations and tensions between these two perspectives, discussing issues such as the validity of rationalistic presuppositions that natural language phenomena are somehow necessarily founded in underlying knowledge structures, and the contrasting validity of naturalistic presuppositions that language games are a human socio-behavioral phenomena arising out of the

dynamics of existence of the species. RIEGER (1991) has recently summarized many of these issues and synthesized them within an integrated framework. CLANCEY (1990) in parallel work concerned with the situated nature of knowledge has drawn attention to the distorted perspectives that derive from the reification of supposed knowledge structures, forming intervening variables in our models of human behavior, as if they actually existed as causal variables within the human systems being modeled.

Many of these issues have a direct practical influence on the methodologies discussed in the following section for eliciting the conceptual structures of psychological individuals (where "eliciting" is itself already a value–laden term). However, it is also useful at this stage to illustrate the relationship between the conceptual structure primitives so far defined and those of lexical semantics (CRUSE 1986) to which they are clearly related.

The arrow of subsumption obviously captures aspects of the taxonomic structure of semantic systems. The top left of Figure 6 shows such a taxonomy (CRUSE 1986:146) in a graphic knowledge representation tool. The finer linguistic distinctions that may be made about such taxonomies, branching versus non–branching, homogeneity in terms of levels, differentiation and convergence at lower levels, and so on, are all additional structural properties of directed acyclic graphs composed from subsumption arrows.

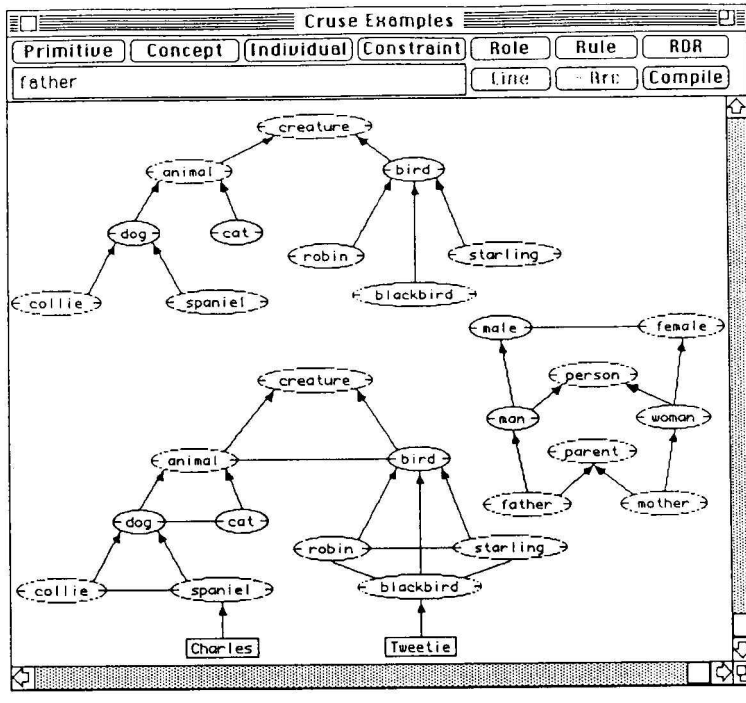


Fig.6 Taxonomies in lexical semantics in a tool based on the visual language

The center right of Figure 6 illustrates how the graphs reflect linguistic semantic arguments. It is based on CRUSE's (1986:262) discussion of *father* and *mother* being "weaker opposites" than *man* and *woman* which, in turn, "are weaker than" *male* and *female*. It can be seen that, for example, *father* and *mother* inherit the disjoint relation between *male* and *female*, but also inherit in common the attributes of *person* and *parent*. Thus, Cruse's analysis of the relative degree of opposition in terms of commonality is apparent in this representation.

At the bottom of Figure 6 the taxonomy in the upper part has been elaborated to show further features of the visual language. For example, *animal* and *bird* have been defined to be opposites in some sense through a disjoint relation. At the very bottom, the representation of concepts intended as rigid designators, or *individuals*, is shown. *Charles* is an instance of a *spaniel*, and *Tweetie* is an instance of a *blackbird*.

It is interesting to consider the relation between Figure 6 and the diagram in Cruse's book. At one level they are isomorphic visual representations of a taxonomic structure. At another level, the computational representation has the added feature that the underlying data structure exists within the computer and hence is readily subject to graph–theoretic analyses of its fine structure. At another level, it is also an operational knowledge structure within the computer supporting automated deductive inference. At another level, this knowledge structure is available for composition with other knowledge structures to provide a knowledge–based system suitable for testing theories and supporting practical applications.

To emphasize the operational nature of Figure 6, Figure 7 shows the KL–ONE–like knowledge structure produced when the "Compile" button is activated in Figure 6.

```
Primitive(creature)
Primitive(animal, creature)
Primitive(animal, bird, creature)
Primitive(dog, animal)
Primitive(dog, cat, animal)
Primitive(collie, dog)
Primitive(robin, bird)
Primitive(collie, spaniel, dog)
Primitive(robin, blackbird, bird)
Primitive(robin, blackbird, starling, bird)
Primitive(male)
Primitive(person)
Primitive(parent)
Primitive(male, female)
Primitive(man, male, person)
Primitive(woman, female, person)
Primitive(father, man, parent)
Primitive(mother, woman, parent)
Individual(Charles, spaniel)
Individual(Tweetie, blackbird)
```

Fig.7 Knowledge structure definitions and assertions compiled from Figure 6

The interpretation of Figure 7 is probably aided by noting that the disjoint relation is specified in this language through a combined syntactic/semantic structure in which predefined concepts in a primitive definition are taken to be the specification of disjoints. Thus,

Primitive(animal, bird, creature)

is unambiguously (but not perspicuously) interpreted as "bird is defined as a kind of creature disjoint from animal."

The knowledge representation server (KRS, GAINES 1991c) inference engine using the structure of Figure 7 to provide a knowledge–based system, would correctly reply in the affirmative to the query "is Charles an animal", and in the negative to "is Tweetie an animal." That is, all the knowledge expressed in Figure 6, and only that knowledge so precisely expressed, would essentially be operationalized through the visual language of Figure 6 being not just diagrammatic but also computational.

In terms of less profound issues, it is significant to note that the expression of knowledge in the visual language has involved the elicitation of not just structure but also terminology. For a single psychological individual, the terminology may regarded as just a lexical reference mechanism. However, across individuals in a community, the determination of whether this reference mechanism is being used uniformly, unambiguously, and without confusion, becomes a very significant aspect of the elicitation of knowledge structures. This issue is a major one in the methodologies described in the following section.

MODELING CONCEPTUAL SYSTEMS

The foregoing section has outlined the theoretical foundations for our studies and the types of model of conceptual structures that we are using. This section addresses the methodological issues of experimentation, data collection and analysis, involved in building such models for actual people and communities. How do the theoretical principles lead to properly corresponding empirical practice?

In restricted, well–defined domains, where knowledge structures are already highly overt, it is possible and attractive to put the graphical tools illustrated in Figure 6 directly into the hands of domain experts and have them enter knowledge structures in the form shown. This is the basis of many graphic knowledge editors in commercial expert system tools and can be a very effective way of making knowledge structures operational. It is now common experience that MacDraw–style graphic tools can be used with great fluency, even by young children, and the visibility of the knowledge structures is helpful in allowing groups to work together to enter shared knowledge.

There is also now a strong track record within the computational linguistics community of tools for manipulating natural language within wide–ranging but restricted domains and constructions, to and from the frame–like structures of Figure 6 (HAUSSER 1987, GOMEZ 1990). Thus, restricted natural language in a technical domain with well–

defined primitive terms is a viable alternative medium for the operational expression of knowledge.

The direct entry of knowledge in what HALL (1959) would term a "technical culture" of highly overt, mutually agreed, definitional knowledge is an important methodology. All the other techniques that we describe, where such direct entry is not possible, or inappropriate to the objectives of a study, ultimately result in such visible overt knowledge structures, and hence they appear as a major end–product of the rationalistic paradigm. There are many methodological issues concerned with the representational schema adopted, the human factors of the interface, and so on, but these will be set aside for the moment to focus on the more profound issues relating to the elicitation of less overt, and more tacit, knowledge. As we have already noted the very vocabulary of the "elicitation" of "tacit" "knowledge" connotes major theoretical presuppositions to which we wish to remain neutral. We shall use this vocabulary because it is standard, but note that we see "elicitation" as a process of self–modeling not presupposing existent knowledge structures, and "tacit knowledge" as a theoretical construct imputed to an agent displaying "intelligent" behavior by an observer.

Extensional Elicitation of Distinctions

The major methodology that we have developed for the elicitation of tacit knowledge is based on extensions of the repertory grid technique originally proposed by KELLY (1955) as an empirical measurement methodology appropriate to personal construct psychology. One can relate the methodology back to the diagram of Figure 3. The conceptual structures in the central region are the imputed basis of the activities of the psychological individuals delimited in the upper part of the diagram. We can attempt to access them by asking the individuals involved to make them overt. However, we can also attempt to access them by asking them to make distinctions about the extensionally defined domain of products in the lower part of the diagram. We can then model the constraints apparent in this extensional data and ascribe these to the imputed conceptual systems in the central section of the diagram. This is the approach taken in repertory grid methodologies.

The psychological basis of the methodology is that people who have difficulty in directly expressing knowledge structures may be able to manifest them in modelable form by applying them to concrete elements within a relevant problem–solving domain. The philosophical basis may be seen as that playing situated language games is an appropriate empirical paradigm in which the imputed knowledge processes may be investigated. The logical basis is that the intensional properties of the subsume and disjoint relations between concepts may be approximated by their extensional properties in terms of the set–theoretic relations between the sub–sets of elements to which the concepts are applied. The approximation arises because of the one way implication between intensional subsumption and extensional inclusion in an open world context, or, in psychological terms, that past behavior is only an indicator, not a determiner, of

future commitments. Repertory grid techniques, illustrated in detail in the next section, elicit knowledge indirectly by prompting individuals for critical elements and relevant constructs in a coherent sub–domain. The constructs define a domain concept that acts as a templet for the cases which are expressed as individuals recognized by the domain concept. Figure 8 shows how the grid relates to the conceptual structure of a domain — a grid is that structure which results from plotting each element in the the extension of a concept against the distinctions that define its intension. Thus, there is a grid corresponding to each domain concept or sub–domain. This corresponds to the application of grid tools in rapid prototyping of knowledge–based systems (GAINES 1988) to characterize a number of significant sub–domains in detail. The significance of grid techniques is that they provide a method of indirectly developing domain concepts through examples when direct methods based on overt knowledge are not available.

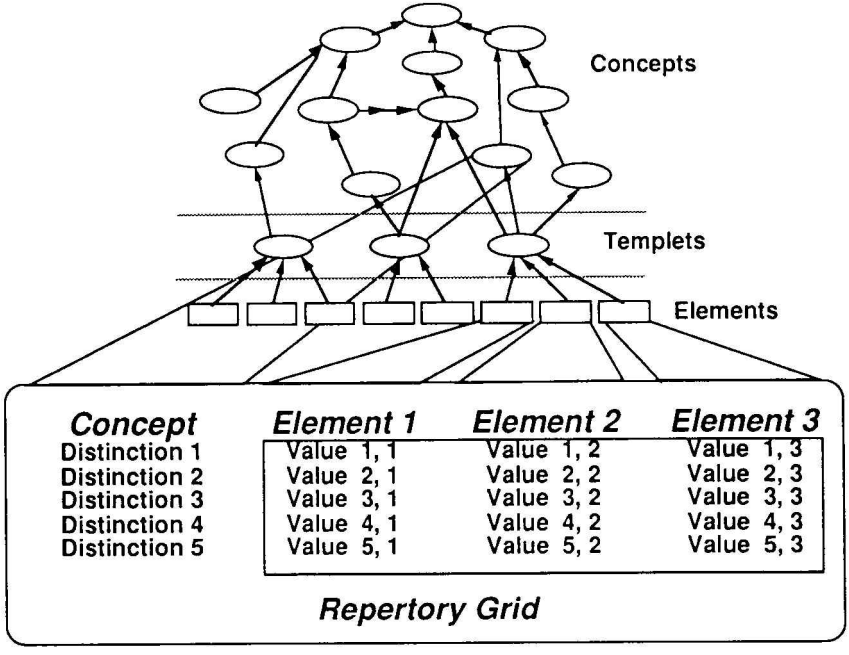


Fig.8 The repertory grid: individuals in the coordinate system of a concept

Repertory grid methodologies are difficult to undertake manually as they require feedback and management from the elicitor while at the same time attempting to avoid inter–personal interactions that would distort the elicitee's conceptual structures. Hence the advent of the personal computer in the mid–1970s and its evolution into the graphic workstations of the 1980s has made the computer implementation of interactive repertory grid elicitation an attractive area of development (SHAW 1980, 1981, MANCUSO; SHAW 1988). This became particularly so in the later 1980s when the need for tools for

interviewing experts in the development of knowledge–based systems became apparent (GAINES; SHAW 1980, SHAW; GAINES 1983, 1987, BOOSE 1984, BOOSE; BRADSHAW 1987, BOOSE; GAINES 1988).

The Comparison of Distinctions

The repertory grid methodology gives a basis for approximating intensional distinctions through their extensions when applied to elements in a domain. The distinctions made by two individuals can then be compared in terms of the differences in their extensions. It is assumed that two distinctions having the same extension may be regarded as evidence that they will also have the same intension. Clearly, this is a fallible hypothesis since increasing the number of elements under consideration may show up differences that were not detected on a smaller set. Hence an important component of repertory grid methodology is feedback on similar distinctions, leading to questions of the form "every element that you distinguish as a you also distinguish as b — can you think of an additional element that is a but not b, or b but not a."

Subject to this limitation, the analysis of extensional distinctions does allow their intensions to be compared. In addition, one may analyze the relations between terminologies, "are two similar distinctions given the same term — is the same term used for two dissimilar distinctions." The two relations of similarity between distinctions and between terminology give rise to a four way classification of concepts as shown in Figures 9 through 12 (GAINES; SHAW 1989, SHAW; GAINES 1989).

In Figure 9, *consensus* arises if the conceptual systems assign the same term to the same distinction.

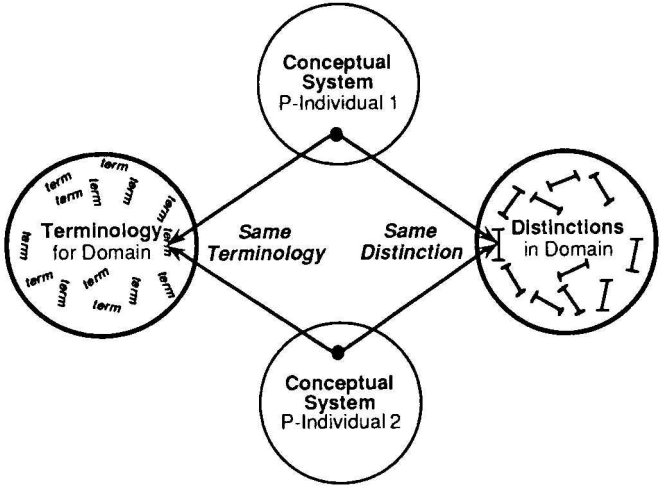


Fig.9 Analyzing the relation of consensus

In Figure 10, *conflict* arises if the conceptual systems assign the same term to different distinctions.

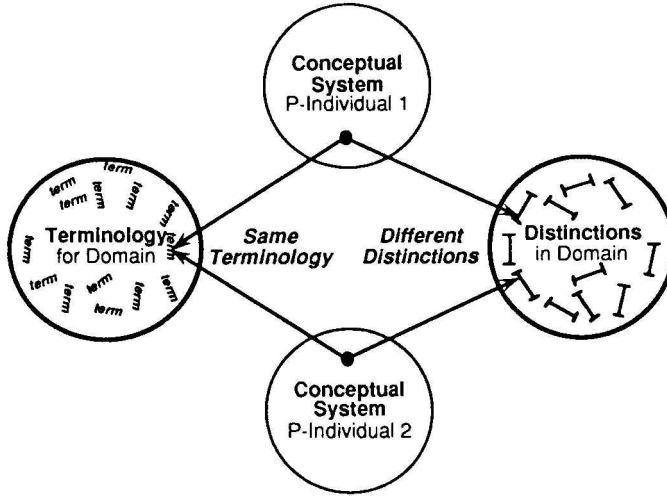


Fig.10 Analyzing the relation of conflict

In Figure 11, *correspondence* arises if the conceptual systems assign different terms to the same distinction.

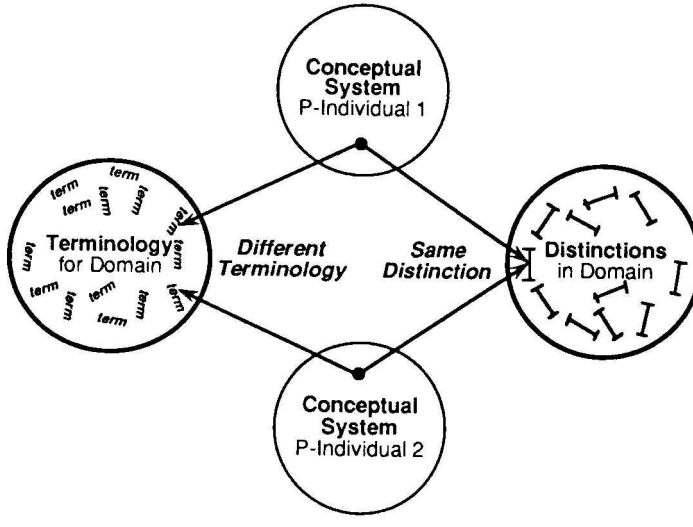


Fig.11 Analyzing the relation of correspondence

In Figure 12, *contrast* arises if the conceptual systems assign different terms to different distinctions.

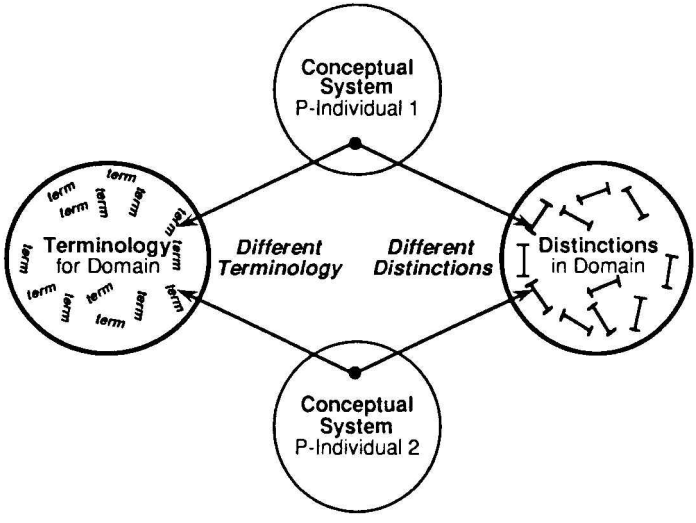


Fig.12 Analyzing the relation of contrast

Figure 13 summarizes this analysis in a four–quadrant diagram that is the basis of feedback presentations to the individuals involved.

Distinctions	Terminology: Same	Terminology: Different
Same	***Consensus*** People use terminology and distinctions in the same way	***Correspondence*** People use different terminology for the same distinctions
Different	***Conflict*** People use same terminology for different distinctions	***Contrast*** People differ in terminology and distinctions

Fig.13 Four–quadrant representation of consensus, correspondence, conflict and contrast between conceptual systems

The recognition of consensual concepts is important because it establishes a basis for communication using shared concepts and terminologies.

The recognition of conflicting concepts is important because it establishes a basis for avoiding confusion over the labeling of differing concepts with same term.

The recognition of corresponding concepts is important because it establishes a basis for mutual understanding of differing terms through the availability of common concepts.

The recognition of contrasting concepts is important because it establishes that there are aspects of the differing knowledge about which communication and understanding may be very difficult, even though this should not lead to confusion. Such contrasts are more common than is generally realized. For example, it is possible to derive the same theorem in mathematics either by using an algebraic perspective, or a geometric one. There is nothing in common in these two approaches except the final result. It may still be possible to discuss the same domain using consensual and corresponding concepts that were not fundamental to the problem solving activities.

TOOLS FOR MODELING CONCEPTUAL SYSTEMS

As already noted, modern personal computers and graphic workstations provide significant capabilities for automating the methodologies described. RepGrid is a tool developed for Macintosh computers that supports the repertory grid elicitation and analysis techniques already described (CPCS 1990).

RepGrid, a Conceptual Elicitation and Analysis Tool

Figure 14 shows the main tools in RepGrid:

- *Elicit* accepts specifications of elements within a domain and provides an interactive graphical elicitation environment within which a person can distinguish elements to derive their constructs. The resultant conceptual system is continuously analyzed to provide feedback prompting the user to enter further elements and constructs. Exchanging grids allows the terms in the conceptual system derived from one group member to be used by another in order to determine whether the two group members have consensus or conflict in their use of terminology and concepts.

- *FOCUS* is a cluster analytic method for the analysis and display of the conceptual systems elicited showing the system as a hierarchical structure.

- *PrinCom* is a cluster analytic method for the analysis and display of the conceptual systems elicited showing the system as a spatial map.

- *Socio* compares elicited and exchanged grids in a variety of ways to determine consensus, conflict, correspondence and contrast. Its methodology is described in detail in the next section.

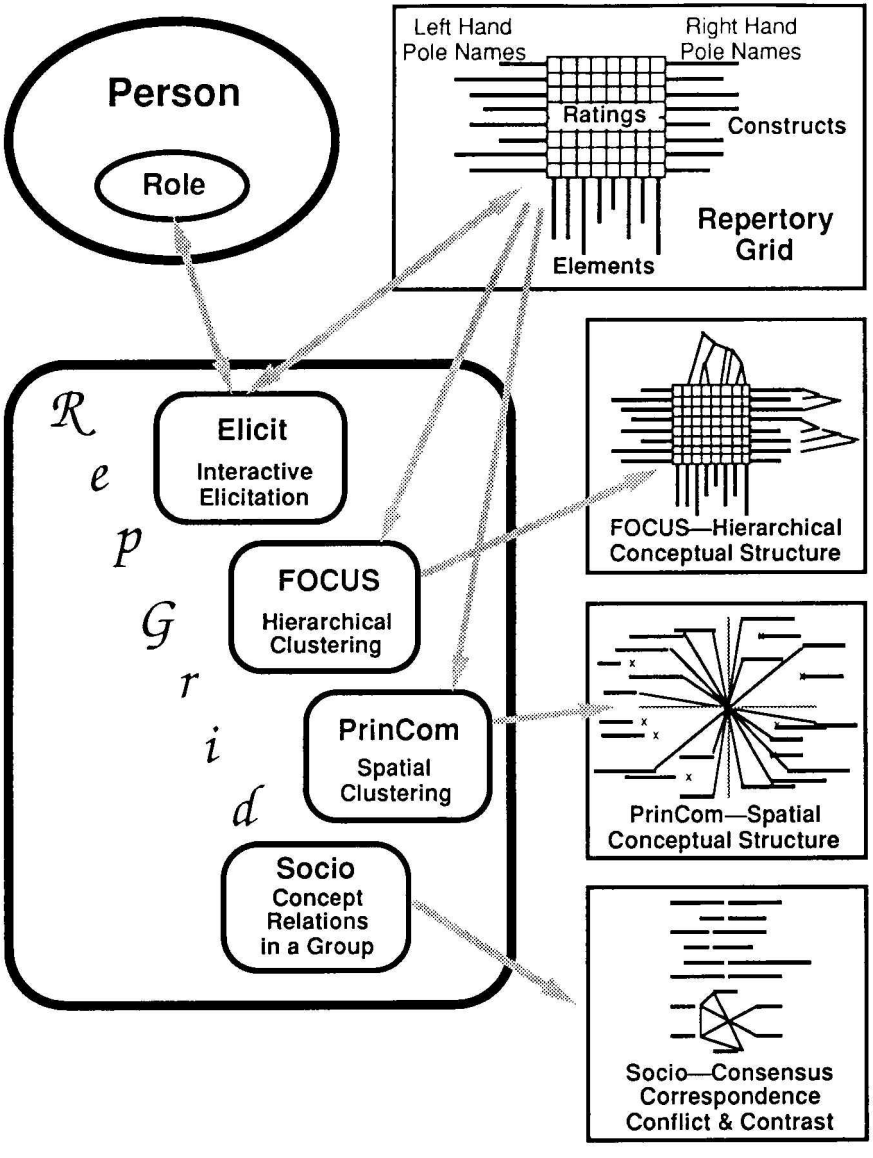


Fig.14 RepGrid tools

RepGrid Elicitation Tools

The next few figures give examples of the interactive graphical elicitation of constructs using RepGrid in a study of the sharing of expertise with multiple experts in a small group of geographers specializing in mapping techniques and their application to geological exploration (SHAW; WOODWARD 1987). They have chosen to refer to the elements as *techniques*, and constructs as *characteristics*. Part of the interaction from the study with one of the geographers is used to give a flavor of *Elicit*. Figure 15 shows the spatial mapping techniques agreed on by all the geographers, and is the start of one of Expert 2's elicitation sessions.

At the top of Figure 15 the buttons show the options that Expert 2 has available on the main element screen. *Triad* and *Pair* are common ways of eliciting constructs by presenting two or three elements respectively, and asking in what way one is different from the other(s). *To Characteristics* changes to the main construct screen, and *Delete, Add*, and *Edit* refer to changing the element set. The bar, and the *Show* button give access to the element matches, and *Status* brings up the screen showing the preliminary set–up variables such as the ratings scale, the grid name etc. Equivalent options are available on the main construct screen.

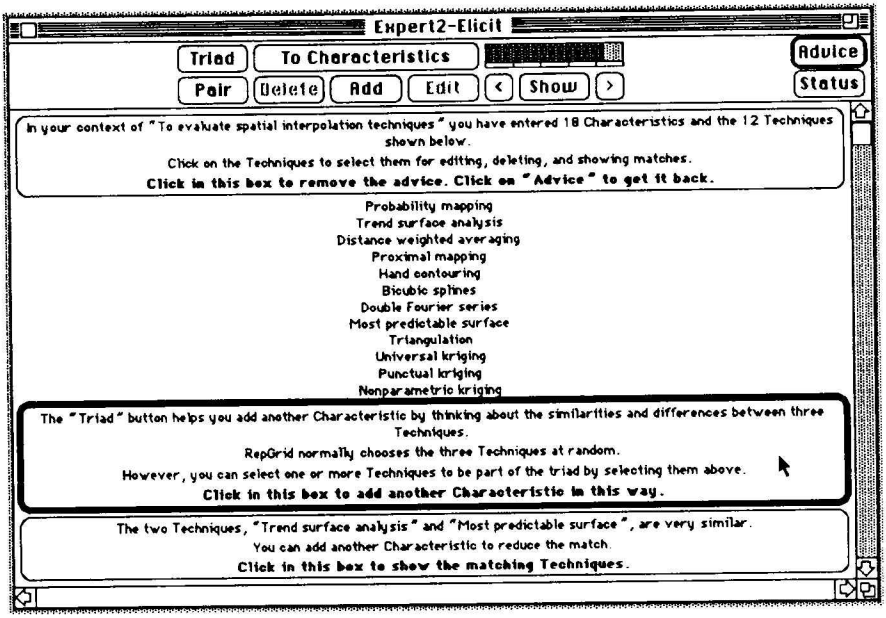


Fig.15 RepGrid screen showing techniques used by the geographers

Expert 2 selects *Triad*, and three techniques are shown as in Figure 16, where she is asked to click the mouse on the one which is different from the other two. She clicks

on *Universal kriging* and elicits the construct *local—global* which is shown in Figure 17 with some techniques already rated, some waiting in a list on the left, and *Proximal mapping* being dragged on to the rating bar. Both the names of the characteristics and the techniques can be edited at any time, and also the placing of the techniques. Figure 18 shows two matched constructs, with both sets of ratings shown, so that they can be compared. Again, any one is moveable if the expert wants to adjust anything.

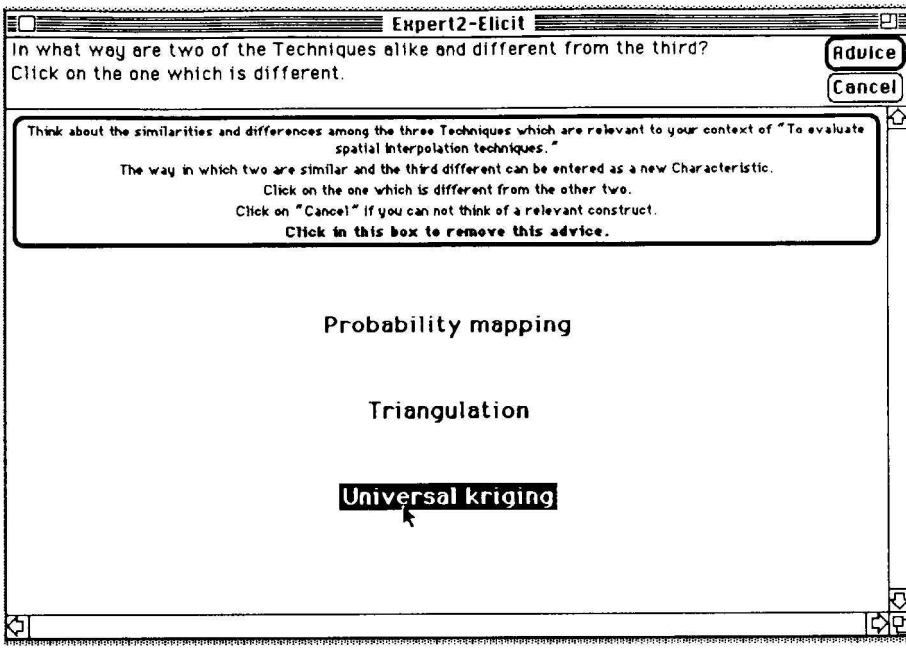


Fig.16 RepGrid construct elicitation from a triad

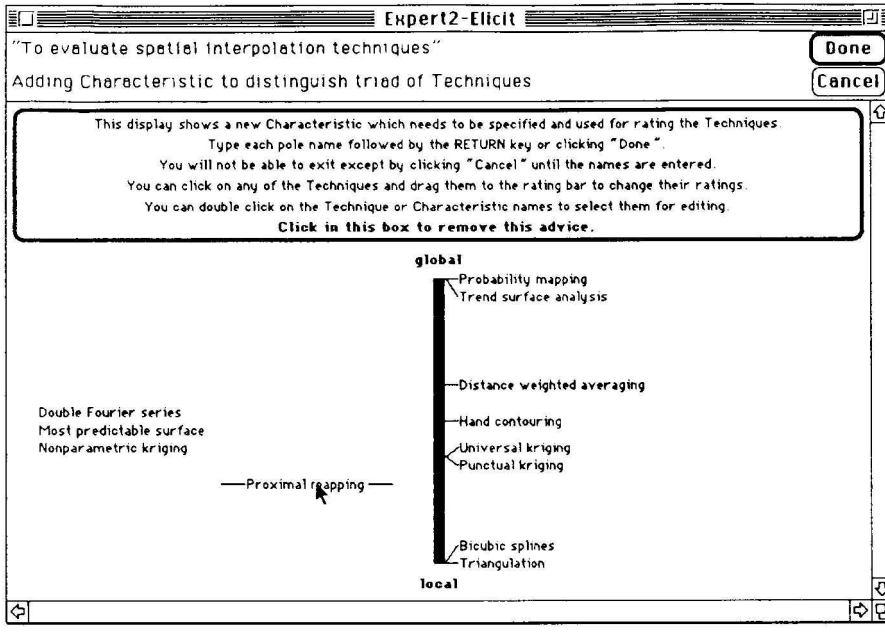


Fig.17 RepGrid showing click and drag elicitation

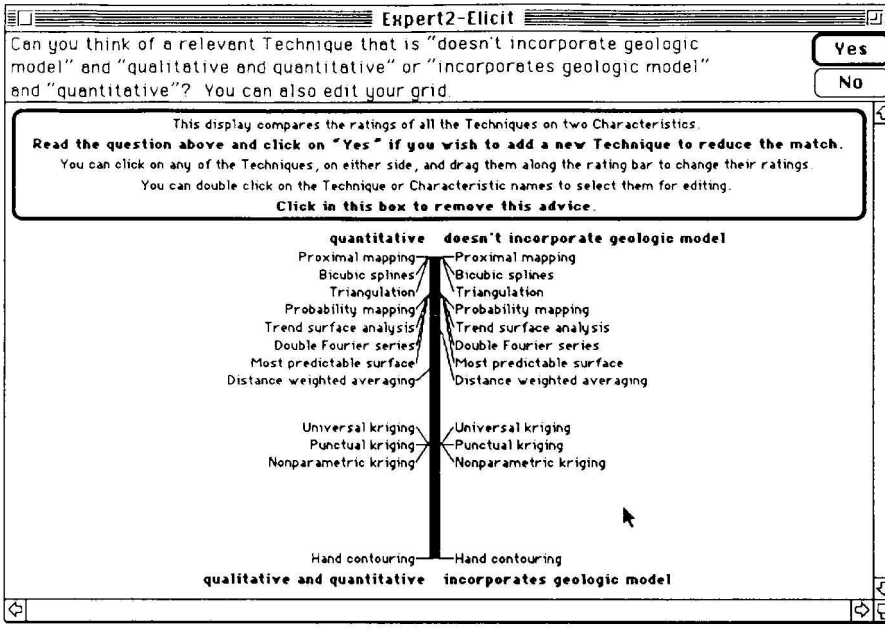


Fig.18 RepGrid showing matching constructs

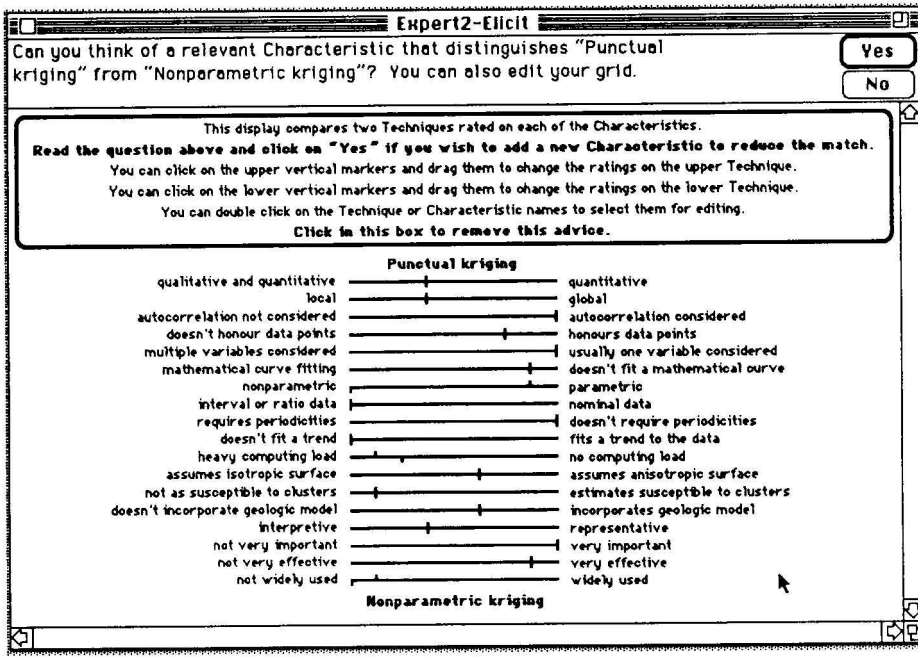


Fig.19 RepGrid showing matching techniques

Techniques may also be matched, as shown in the screen in Figure 19. The technique at the top has a marker on each construct pointing up, and the one at the bottom a marker pointing down. Again, any marker is moveable if the expert wants to change a rating. The *Advice* button at the top right of the main element and construct screens will automatically provide the best option at any stage of the grid elicitation if the user is unsure which to choose.

RepGrid Individual Grid Analysis Tools

Figure 20 displays the final grid. The columns represent the ratings of the elements listed at the bottom of the display, and the rows the constructs. In order to show where each element is placed on each construct, a number has been used, in this case 1 to 9 — 1 representing the left pole label, and 9 the right pole label. Hence *Probability mapping* is rated 9 on *local—global*, indicating that it is global, whereas *Bicubic splines* is rated 1, local.

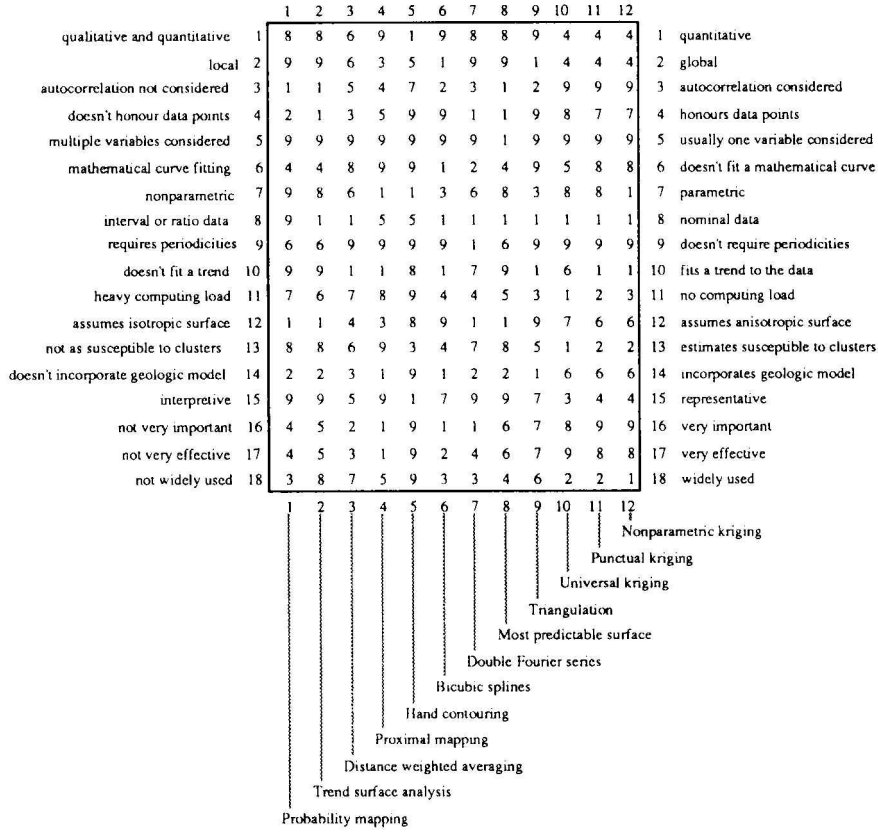


Fig.20 RepGrid display of completed grid

Figure 21 shows the grid clustered by the FOCUS algorithm. This involves the rows and columns being reordered to put the ones most like each other close together. As can be seen on the construct tree at the top right, there are three clusters at 75%, and one construct which is not part of any cluster. The bottom right tree of techniques has been clustered to the 70% level. From this, it can be seen that the one most different from the others is *Hand contouring*. The others are divided into two main clusters. The rating values have been shaded to make it easier to distinguish the blocks on the left pole from those on the right. Many other conclusions can be drawn from this diagram, but it is sufficient here just to give an example of what the expert experienced in using RepGrid.

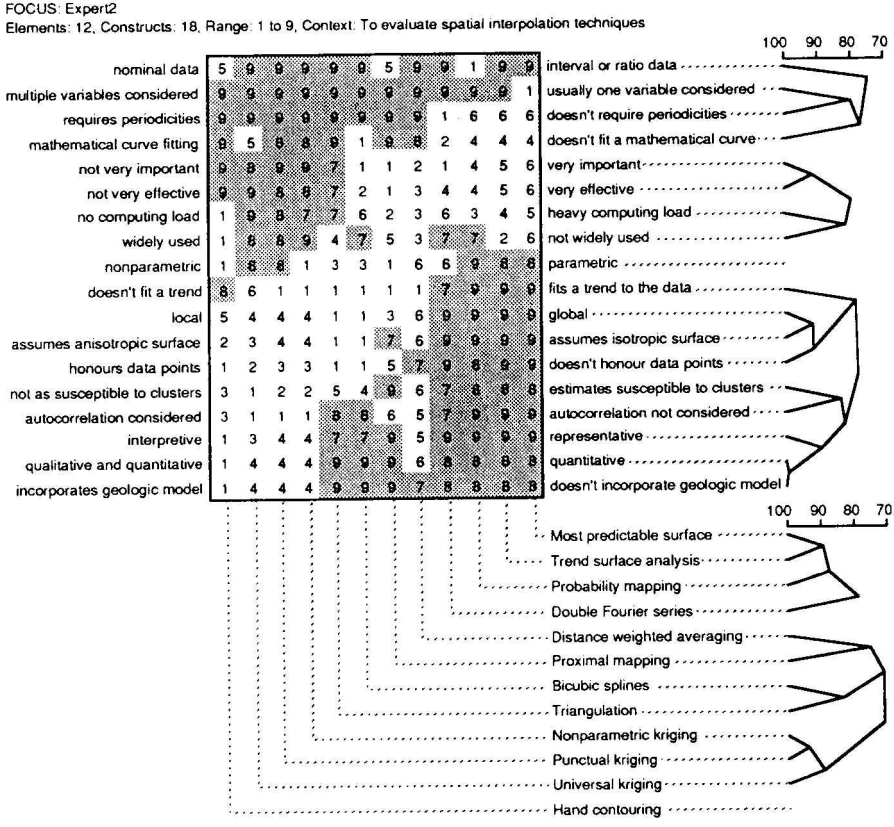


Fig.21 RepGrid FOCUS of completed grid

Figure 22 shows the output from PrinCom, the principal component analysis of the grid. Here the elements (in italic) and constructs are plotted in the same space, showing in a different way from FOCUS how they are related. However, the main clusters are apparent in either diagram, the main differences being where the least matched elements and constructs are placed.

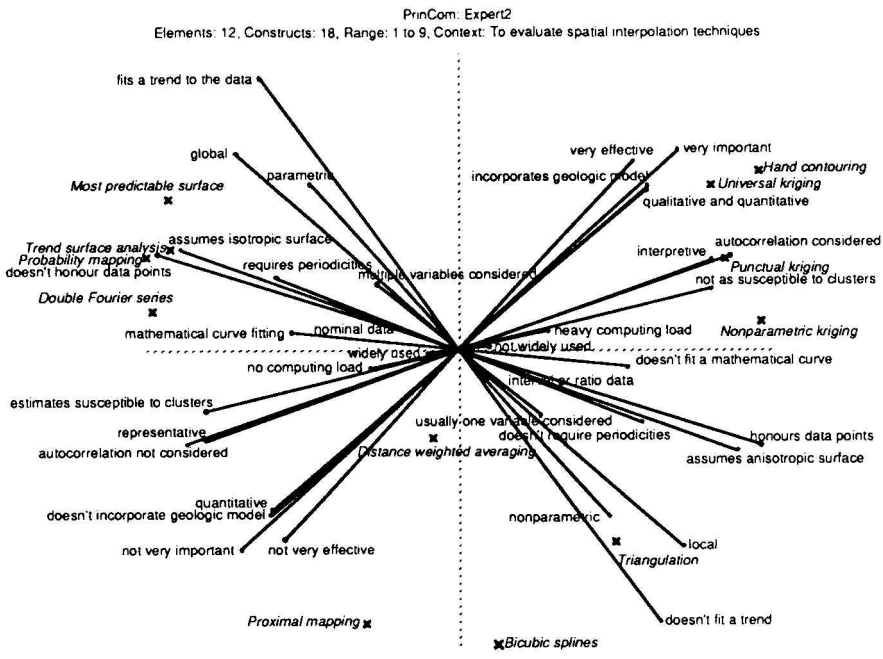


Fig.22 RepGrid PrinCom of completed grid

RepGrid Multiple Grid Analysis Tools

Exchange methodologies were developed for the measurement of understanding and agreement between either two individuals, two roles or on two occasions (SHAW 1980). To do this two group members, possibly experts with differing points of view, each elicit a grid in an area of common knowledge or experience. Each may choose his own elements independently of the other, and elicit and rate his or her constructs quite separately. Each then can *Exchange* his or her grid, that is use the other's elements and constructs but fill in his or her own rating values. For example, in terms of Figure 17, the exchanging expert would see the terms *local—global*, but the elements would all be to the left and have to be dragged to the scale with no knowledge of where the other expert had previously placed them.

The *Socio* analysis in RepGrid allows people who are members of a community of discourse to explore their agreement and understanding with other members, and to make overt the knowledge network involved (SHAW 1980, 1981, 1988). It is an extension of techniques such as SOCIOGRIDS (SHAW 1980) for deriving socionets and mode constructs from groups of individuals construing the same class of elements. The objective of Socio is to take different conceptual systems in the same domain and

compare them for their structure, showing the similarities and differences. It may be regarded as the implementation of a simple form of analogical reasoning. Figure 23 shows the basis of operation of Socio — consider one set of data as being the base class defined by its elements, their constructs and values, and consider variant classes:

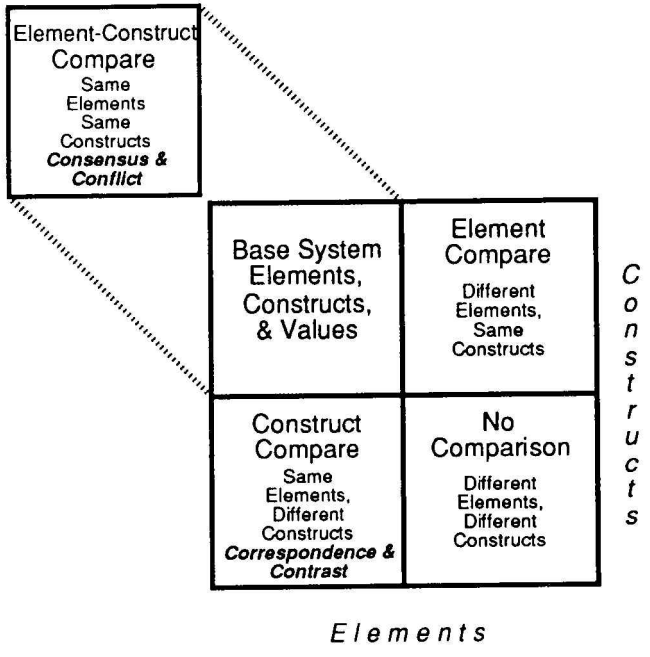


Fig.23 Possible comparisons between base and related systems

- *Element–Construct Compare*: The conceptual system at the upper left has the same elements and constructs but possibly differing values. Socio analyzes the matches between the elements and the constructs in this and the base class according to the values, and shows those elements and constructs that are similar and those which are different. A typical application is to see whether group members agree on the definitions of constructs by asking them to separately fill in the values for a grid exchanged between them.

- *Element Compare:* The conceptual system at the upper right has different elements but the same constructs. Socio analyzes the matches between the elements in this and the base class, and for each element in the base class shows the closest matching element in the other class. A typical application is to see whether group members are using different terminologies for the same elements by asking them separately to define elements and fill in the values for a domain defined through a class with agreed constructs.

- *Construct Compare:* The conceptual system at the lower left has different constructs but the same elements. Socio analyzes the matches between the constructs in this and the base class, and for each construct in the base class shows the closest matching constructs in the other class. A typical application is to see whether group members are using different terminologies for the same constructs by asking them to separately define constructs and fill in the values for a sub–domain defined through a class with agreed elements.

If, as in the lower right, neither elements nor constructs are common then no comparison is possible.

When a number of classes representing the same sub–domain are available, Socio also provides two other forms of analysis:

- *Mode Elements and Constructs:* Socio attempts to derive "modal" elements or constructs that reflect a consensus among the group members. It does this by extracting those which occur as highly matched elements or constructs across the majority of conceptual systems. A typical application is to reach consensus on critical concepts that are associated with a rich vocabulary at differing levels of abstraction.

- *Socionets:* Socio derives a socionet showing the degree to which each member is able to make the same distinctions as another, even if they use different terminology.

A METHODOLOGY TO DERIVE CONCEPTUAL RELATIONS

Using the concepts developed above it is possible to develop a complete methodology for eliciting and analyzing consensus, conflict, correspondence and contrast in a group, and implement this as an automatic process using the tools in RepGrid. The methodology has three phases shown in Figures 24 through 26.

Phase 1: Domain Discussion and Instantiation

In phase 1 people who are members of a community of discourse come to an agreement over a set of elements which instantiate the relevant domain. This is the initial phase of any repertory grid methodology, whether used with individuals or groups. However, with individuals the elicitation techniques may be used to elicit more elements as the exploration of the conceptual domain proceeds. When comparing group members it is important that a set of elements is established at the start of the comparative study, and that they mutually agree on the definitions of these.

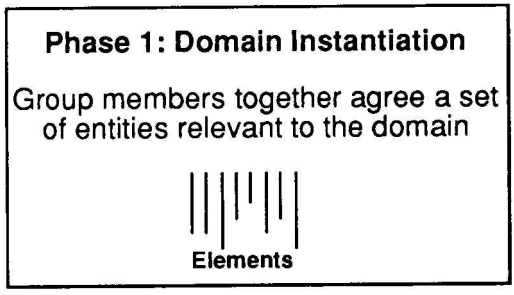


Fig.24 Methodology Phase 1

A convenient way to generate this set of elements is have each group member individually use Elicit to enter his or her conceptual system for a domain, and then extract the elicited elements from all the grids for discussion and consolidation by the group. This has the advantage that the group members gain some experience in the use of the RepGrid tools and can take advantage of the full elicitation facilities.

Phase 2: Conceptualization and Feedback

In phase 2 each group member individually elicits constructs and values for the agreed elements. The resultant conceptual systems will have the same elements but different constructs and can be analyzed by the *Construct Compare* component of Socio as shown in Figure 23. This takes each construct in one grid and determines the best matching construct in the other grid, if there is one. The result is a mapping from the constructs in one group member's grid to those in another's as shown in Figure 25.

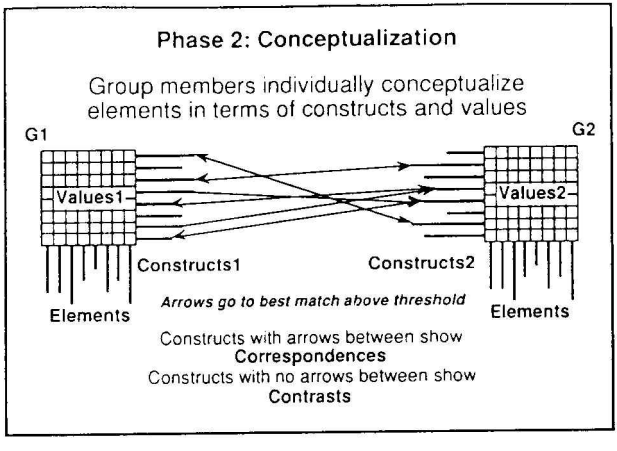


Fig.25 Methodology Phase 2

In evaluating this mapping we are not particularly interested in the terminology used but rather whether one group member has a construct that can be used to make the same distinctions between the elements as does the other group member, regardless of whether these distinctions are called by the same terms. If such a *correspondence* occurs then the group members have a basis for mutual understanding of the underlying concept.

If a construct in one system has no matching construct in the other then it stands in *contrast* to all the other group member's constructs, and it may be very difficult for the other group member to understand the use of this term.

Note that the arrows in Figure 25 need not be symmetric. Construct A in G1 may be best matched by construct B in G2, but construct B in G2 may be best matched by construct C in G1. The only constraint is that if an construct has an incoming arrow then it will have an outgoing arrow.

Thus, the second phase provides the basis for analysis of correspondence and contrast as already discussed.

Phase 3: Exchange and Compare

In phase 3 each group member individually exchanges elicited conceptual systems with every other group member, and fills in the values for the agreed elements on the constructs used by the other group members. The resultant conceptual systems will have the same elements and constructs and can be analyzed by the *Element–Construct Compare* component of Socio as shown in Figure 23. This takes each construct in one grid and determines whether it matches the corresponding construct in the other grid.

The result is a map showing *consensus* when constructs with the same labels are used in the same way and *conflict* when they are not as shown in Figure 26. Thus, the third phase provides the basis for analysis of consensus and conflict as already discussed.

Figure 26 also shows the correspondence and contrast relations analyzed in phase 2 as relations between two of the grids used in phase 3. Thus, for two group members, four gri ds obtained by one elicitation and one exchange each, are sufficient to classify the relations between constructs in terms of consensus, conflict, correspondence and contrast. The methodology s cales up linearly for each group member, so that n group members will be involved in n elicitations, one base elicitation and $n - 1$ exchanges.

These three phases result in the group members' conceptual systems having become overt and inter–related. They lead naturally to later phases in which classes, objects and rules can be developed incorporating consensual, corresponding, and some of the contrasting constructs as kernel knowledge and, possibly, the conflicting and remaining contrasting constructs as 'other opinions.'

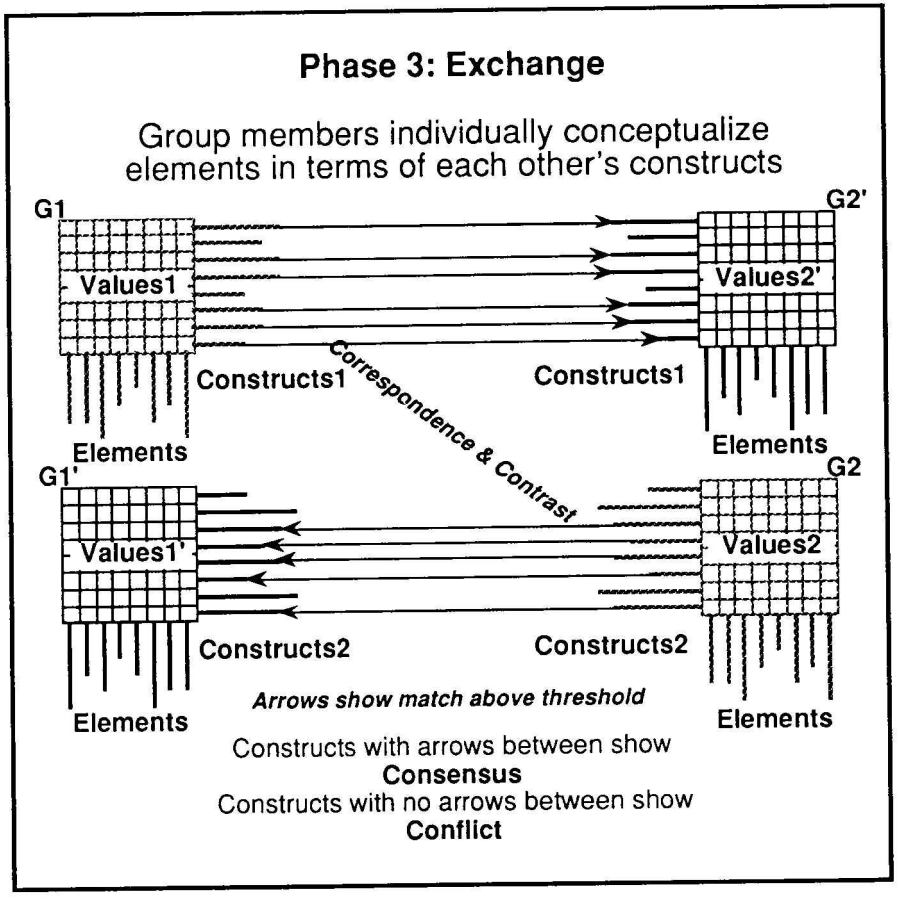


Fig.26 Methodology Phase 3

An Example of the Methodology in Action

This section gives an example of the methodology in action using data from the study of geographers specializing in mapping techniques previously cited (SHAW; WOODWARD 1987).

G1:G2 33.3% over 80.0 (ExpertB construct–consistency–with ExpertA)

1:	8.3% ≥	90.9	A2:	Interval data — Nominal data
2:		81.8	A4:	Global — Local
3:		81.8	A5:	Intuitive — Mathematical
4:	33.3% ≥	81.8	A6:	Requires spatial search — Does not require spatial search
5:	41.7% ≥	75.0	A10:	Difficult to understand — Easily understood
6:		72.7	A3:	Non–polynomial — Polynomial
7:	58.3% ≥	72.7	A7:	Discontinuous — Continuous
8:	66.7% ≥	59.1	A12:	Does not consider non–spatial attributes — Considers non–spatial attributes
9:	75.0% ≥	56.8	A11:	Few points — Many points
10:	83.3% ≥	50.0	A8:	Does not honour data — Honours data
11:		47.7	A1:	Requires no model — Requires model
12:	100.0% ≥	47.7	A9:	Linear interpolation — Non–linear interpolation

G1:G2 8.3% over 80.0 (ExpertB element–consistency–with ExpertA)

1:	9.1% ≥	83.3	E11:	Vector trend surface analysis
2:	18.2% ≥	77.1	E2:	Trend surface analysis
3:		72.9	E4:	Distance weighted averaging
4:	36.4% ≥	72.9	E7:	Bicubic splines
5:		70.8	E3:	Kriging
6:		70.8	E5:	Proximal mapping
7:		70.8	E8:	Double Fourier series
8:	72.7% ≥	70.8	E9:	Most predictable surface
9:	81.8% ≥	60.4	E10:	Negative exponential surface
10:	90.9% ≥	56.2	E6:	Hand contouring
11:	100.0% ≥	43.8	E1:	Probability mapping

Fig.27 Element–construct comparison of expert B with expert A

Figure 27 shows an element–construct comparison from the geographic study in which expert B has filled in the values for a class defined by elements and constructs elicited from expert A. The print out shows the matches sorted with best first. The cumulative percentage is given of the number of constructs with matches greater than the value shown. In the list of constructs at the top, it can be seen that there is consensus on *interval data—nominal data*, but conflict on *requires no model—requires a model* and *linear interpolation—nonlinear interpolation*. There is clear consensus on the top four constructs, clear conflict on the lower five, and uncertainty about the remaining three.

In the list of elements at the bottom, it can be seen that there is close agreement on *vector trend analysis*, but high disagreement on *probability mapping*. This output can be used to focus a discussion between the experts on why they differ in their views of *probability mapping* and the classification of mapping techniques in terms of *linear* or *nonlinear interpolation*. For example, the first construct, *requires no model—requires a model* shows high disagreement, and looking further into the elicited data it can be seen that expert A thinks that *probability mapping requires a model*, whereas expert B thinks that *probability mapping requires NO model*. On inquiring into this, the explanation

given was couched in terms of what one actually means by the term "model", indicating the conflicting use of terminology.

Figure 27 can be redrawn as a *difference grid* where rating values (in this case 1 to 5) for expert B's ratings of expert A's elements on his constructs are subtracted from expert A's similar rating values respectively. Figure 28 shows this with the elements and constructs about which they agree the most in the top right corner, shown by no difference or a difference of only 1; and those with most disagreement towards the bottom left, shown by the maximum difference of 4 or a large difference of 3. The darker gray areas indicate the greater difference. The graphs on the right show the declining matches as in Figure 27, and may be used to decide where to place thresholds distinguishing consensus and conflict. Hence from this difference grid, the consensus and conflicts can easily be identified and discussed by the experts.

The presentation of Figure 28 contains sufficient raw and processed data that it may be used for a range of discussions, from identification of apparent disagreements, such as that on the use of the term "linear interpolation" and the entity "probability mapping," to a detailed analysis of how these disagreements have arisen. Measures of correspondence and contrast shown in Figure 29 are presented in a similar graphical form with distinctions paired from the two grids.

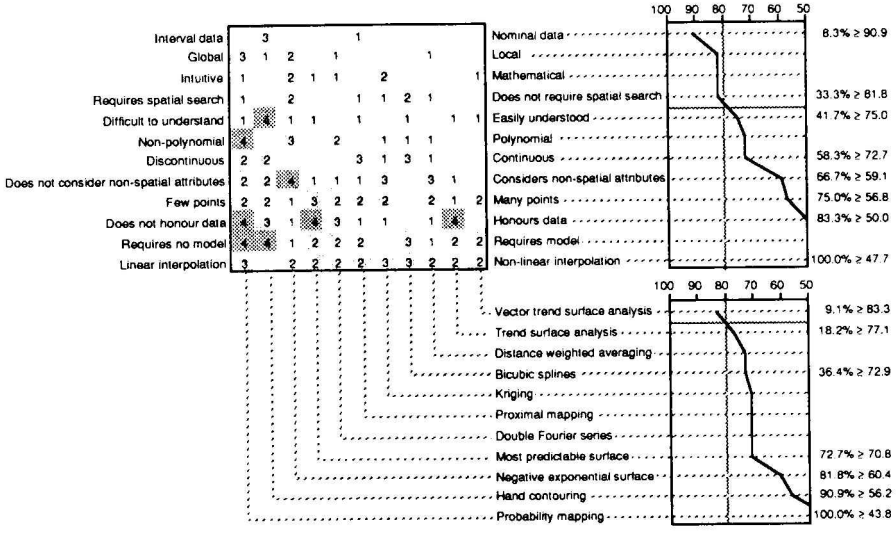


Fig.28 The difference grid for experts A and B

G1<:G2 62.5% over 80.0 (ExpertA construct–construed–by ExpertB)

1:	6.2% ≥ 88.5	G1A2:	Local — Global
		G2A2:	local — global
2:	12.5% ≥ 87.5	G1A3:	Low level data — High level data
		G2A8:	nominal data — interval or ratio data
3:	86.5	G1A1:	Does not honour data points — Honours data points
		G2A4:	doesn't honour data points — honours data points
4:	86.5	G1A7:	Short distance autocorrelation — Long distance autocorrelation
		G2A2:	local — global
5:	31.2% ≥ 86.5	G1A9:	New geographical technique — Old geographical technique
		G2A18:	not widely used — widely used
6:	37.5% ≥ 85.4	G1A16:	Not widely used — Widely used
		G2A18:	not widely used — widely used
7:	43.8% ≥ 83.3	G1A5:	Discontinuous — Continuous
		G2A2:	local — global
8:	50.0% ≥ 82.3	G1A4:	Mathematically complex — Mathematically simple
		G2A11:	heavy computing load — no computing load
9:	81.2	G1A10:	Hard to adapt to multivariate — Easy to adapt to multivariate
		G2A5:	usually one variable considered — multiple variables considered
10:	62.5% ≥ 81.2	G1A12:	Does not require spatial search — Requires spatial search
		G2A13:	estimates susceptible to clusters — not as susceptible to clusters
11:	79.2	G1A6:	Does not require a priori model — Requires a priori model
		G2A2:	local — global
12:	75.0% ≥ 79.2	G1A11:	Few points — Many points
		G2A18:	not widely used — widely used
13:	76.0	G1A13:	Does not use polynomial — Uses polynomial
		G2A6:	doesn't fit a mathematical curve — mathematical curve fitting
14:	87.5% ≥ 76.0	G1A15:	Not very effective — Very effective
		G2A17:	not very effective — very effective
15:	93.8% ≥ 72.9	G1A14:	Not very important — Very important
		G2A18:	not widely used — widely used
16:	100.0% ≥ 71.9	G1A8:	Models the stationarity — Assumes stationarity
		G2A3:	autocorrelation not considered — autocorrelation considered

Fig.29 Construct comparison of expert B with expert A

Figure 29 shows a construct comparison from the geographic study in which expert B has specified constructs and filled in the values for a class defined by elements elicited from expert A. The print out shows the matches sorted with best first. The cumulative percentage is given of those with matches greater than the value shown, and the construct from B which best matches each from A is shown beneath it. It can be seen that:

- The first or highest match which accounts for 6.2% of the constructs has a level of 88.5 out of a possible 100 if they were identical. That is, both experts are using the construct *local—global* in the same way. This is an example of *correspondence*. However, it is not an interesting one because we can see that the terms used are the same and it is effectively arising from a *consensus*.

- The first and second matches together account for 12.5% of all constructs, and they are matched over the level of 87.5. For the second construct, when Expert A uses the term *low–level data—high–level data*, expert B is using the term *nominal data—interval or ratio data*. This shows a difference in terminology which can be interpreted as their *levels of abstraction* being different in their construing of this topic. This illustrates the two experts having constructs in *correspondence*.

- The third match again shows a *correspondence* that can be interpreted as arising from *consensus*, with both experts using the construct *does not honour data points—honours data points* in the same way. Note that this showed up as *conflict* in Figure 27, probably reflecting that, in this study, the methodology was more complex and the exchange grids from which Figure 27 derives were elicited and discussed *before* those from which Figure 29 derives.

- The fourth match shows a *correspondence* between *short–distance autocorrelation—long–distance autocorrelation* and *local—global*. Notice that *local—global* used by expert B was also used in the first match indicating that expert A has two constructs *short–distance autocorrelation—long–distance autocorrelation* and *local—global* which are used similarly to each other but with different terminology, whereas expert B has only one. In fact, looking on to the seventh and eleventh matches, it can be seen that expert A has two more constructs *discontinuous—continuous* and *does not require a priori model—requires a priori model* which correspond to expert B's single construct *local—global*. This shows a difference in richness of concepts not necessarily making new distinctions in the class so far defined by the elements.

- The eighth match, still over the level of 82, again shows *correspondence*. It shows the construct *heavy computing load—no computing load* is being used by expert B to correspond to *mathematically complex—mathematically simple* used by expert A. We can interpret this as a difference in terminology corresponding to a correlation in the real world.

Figure 30 shows each of the constructs from Figures 27 and 29 put into the appropriate quadrant of Figure 13. There is no significant example of *contrast* in this data, possibly because the two experts work very closely together. Such examples do arise in a full analysis of the three experts in the original study.

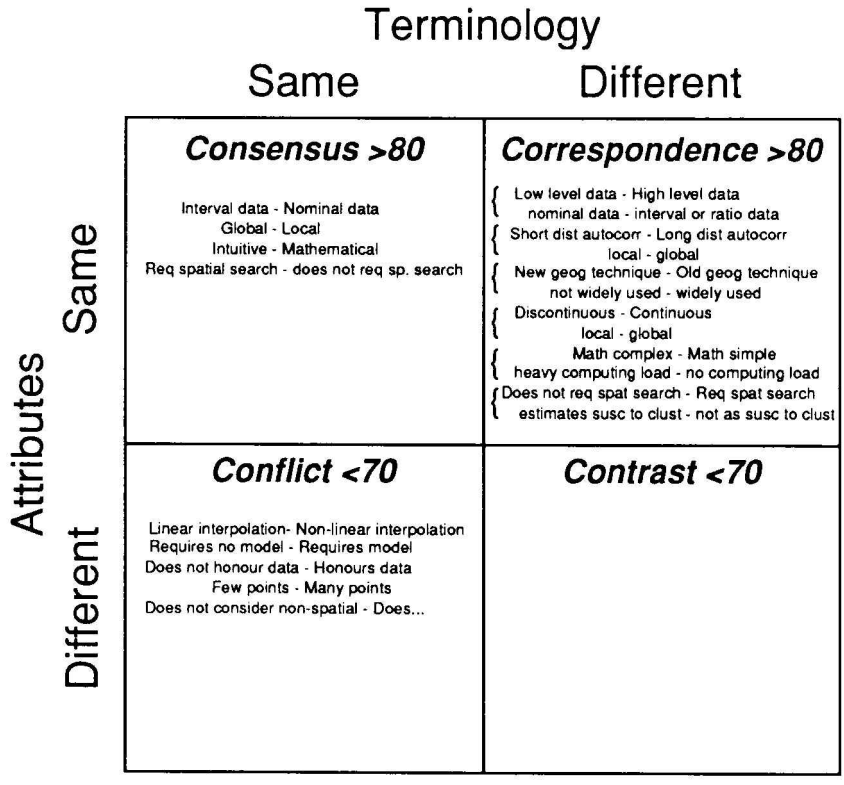


Fig.30 Consensus, correspondence, conflict and contrast from Figs.27 and 29

Figure 31 shows a *mode constructs* analysis of data from the three geographers which extends the clusters noted above. These mode constructs are essentially inter–conceptual–system clusters of corresponding constructs from the three experts. The first may be interpreted as centering around *Global–Local* encompassing a variety of concepts related to this; the second as around autocorrelation techniques and their consequences; the third as around the complexity of the technique; the fourth around the type of data; and the fifth around the number of variables considered.

Mode Construct 1: 13 constructs in 3 grids at 80.0

G1A1:	Does not honour data points	Honours data points
G1A2:	Global	Local
G1A5:	Continuous	Discontinuous
G1A7:	Long distance autocorrelation	Short distance autocorrelation
G2A2:	global	local
G2A4:	doesn't honour data points	honours data points
G2A7:	parametric	nonparametric
G2A9:	requires periodicities	doesn't require periodicities
G2A10:	fits a trend to the data	doesn't fit a trend
G2A12:	assumes isotropic surface	assumes anisotropic surface
G3A1:	global	local
G3A6:	periodicity	non–periodicity
G3A7:	very smooth	non–smooth data

Mode Construct 2: 8 constructs in 3 grids at 80.0

G1A12:	Does not require spatial search	Requires spatial search
G2A3:	autocorrelation not considered	autocorrelation considered
G2A13:	estimates susceptible to clusters	not as susceptible to clusters
G2A15:	representative	interpretive
G3A3:	data restrictions	no data restrictions
G3A8:	low computer cost	high computer cost
G3A9:	no error estimates	error estimates
G3A13:	not so effective technique	very effective technique

Mode Construct 3: 6 constructs in 3 grids at 80.0

G1A4:	Mathematically complex	Mathematically simple
G1A9:	New geographical technique	Old geographical technique
G1A16:	Not widely used	Widely used
G2A11:	heavy computing load	no computing load
G2A18:	not widely used	widely used
G3A11:	non–linear surface	linear surface

Mode Construct 4: 4 constructs in 3 grids at 80.0

G1A3:	Low level data	High level data
G2A8:	nominal data	interval or ratio data
G3A2:	nominal	interval
G3A5:	non–continuous	continuous

Mode Construct 5: 2 constructs in 2 grids at 80.0

G1A10:	Hard to adapt to multivariate	Easy to adapt to multivariate
G2A5:	usually one variable considered	multiple variables considered

Fig.31 Mode constructs from three experts

These five mode constructs can be interpreted as indicating stereotypical lines of reasoning most used by these experts. This output can be used as a basis for discussion among the experts on whether these conceptual clusters should be split because they confound different concepts expressed in apparently corresponding constructs, or retained as being the same concept expressed in different terminologies. Once this form of analysis has been discussed by the group it is readily edited and extended.

Construct Links

G3<:G2 71.4% over 80 (ExpertC construct–construed–by ExpertB)
G1<:G2 62.5% over 80 (ExpertA construct–construed–by ExpertB)
G2<:G1 61.1% over 80 (ExpertB construct–construed–by ExpertA)
G2<:G3 44.4% over 80 (ExpertB construct–construed–by ExpertC)
G3<:G1 42.9% over 80 (ExpertC construct–construed–by ExpertA)
G1<:G3 31.3% over 80 (ExpertA construct–construed–by ExpertC)

Fig.32 Socionet analysis of three experts

Figure 32 shows a *socionets* analysis of the same data based on a set of comparisons like that shown in full in Figure 29. It can be seen from the first two links that the conceptual system of expert B encompasses the majority of constructs used by experts A and C. This indicates that expert B has a deeper knowledge of the topic than either expert A or expert C. That of expert A encompasses the majority of the constructs of B, shown in the third link. However, that of expert C does not encompass many of those of A and B, and that of A does not encompass many of those of C indicating that C has a different point of view, or a background of different experiences from those of A and B. This could then be explored in detail with expert C.

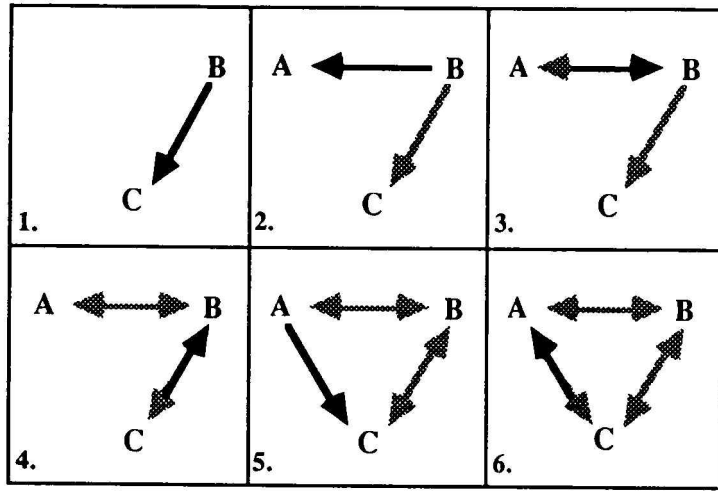


Fig.33 Socionet of relations between experts' conceptual systems

Figure 33 shows this information expressed in the socionet links of relations between the experts' conceptual systems for the domain. Each new link is shown by a black arrow as it is added into the sequence.

RECENT DEVELOPMENTS

This section provides a brief overview of recent developments in our conceptual modeling tools which are described in detail elsewhere. The most notable are: experiments in moving the tools described in this paper out into the computer–networked community to become a routine service as an extension to electronic mail; developments of integrated knowledge support system architectures combining conceptual elicitation, visual languages, machine induction and hypermedia techniques.

RepGrid–Net (SHAW; GAINES 1991a,b) is an attempt to move computer–based techniques for supporting the analysis of group cognitive processes and decision–making from being specialist applications, typically used once in studying or supporting a group, to becoming a routine organizational tool used as readily as electronic mail. We have taken electronic mail as our starting point for a number of reasons. First, it is already widely used within organizations on a routine basis (KERR; HILTZ 1982, QUARTERMAIN 1990). A logical, and psychologically harmonious, extension to electronic mail is far more likely to be accepted and used routinely than a new application where a major decision has to be made about use. Second, we see human discourse as an essential process in the formation of functional groups within society that is supported at a very basic level by electronic mail. Our objective is to extend the support of discourse through tools that make conceptual models more overt as an aid to mutual understanding — what HABERMAS (1981) has termed "communicative action."

RepGrid–Net is a computer–based message system that integrates conventional electronic mail and bulletin board facilities with the repertory grid elicitation and analysis facilities described above to provide both unstructured and structured communications supporting the formation and operation of special interest networks. Users see a mail system in which special–interest groups are specifically supported. The coordinator of a special–interest group provides a basic focus for it through statements of intent, topics and issues which are handled on a bulletin board basis. He or she also provides one or more kernel grids listing specific topics and the concepts which they apply to them. These kernel grids can be developed by others interested in the groups, using the stated topics and concepts, and adding to them. General similarities between grids are analyzed to provide a socionet of people with common viewpoints, and this may be used to access the mail system to communicate with them. Detailed comparisons of similarities and differences between viewpoints may be made, and individual concept structures can be analyzed.

Figure 34 shows the overall systems architecture. The message sub–system is conventional in its operation. Users register with the system and may access lists of those with mailboxes registered on the system, lists of incoming mail and confirmations that outgoing mail has been accessed. They may prepare and send mail, and receive and reply to mail. The special interest group handles the group bulletin board and grid elicitation and analysis. The integration sub–system manages the interactions between the other two sub–systems, for example, that the graphic displays of networks of people with common interests may be used to send a message to one or more of them.

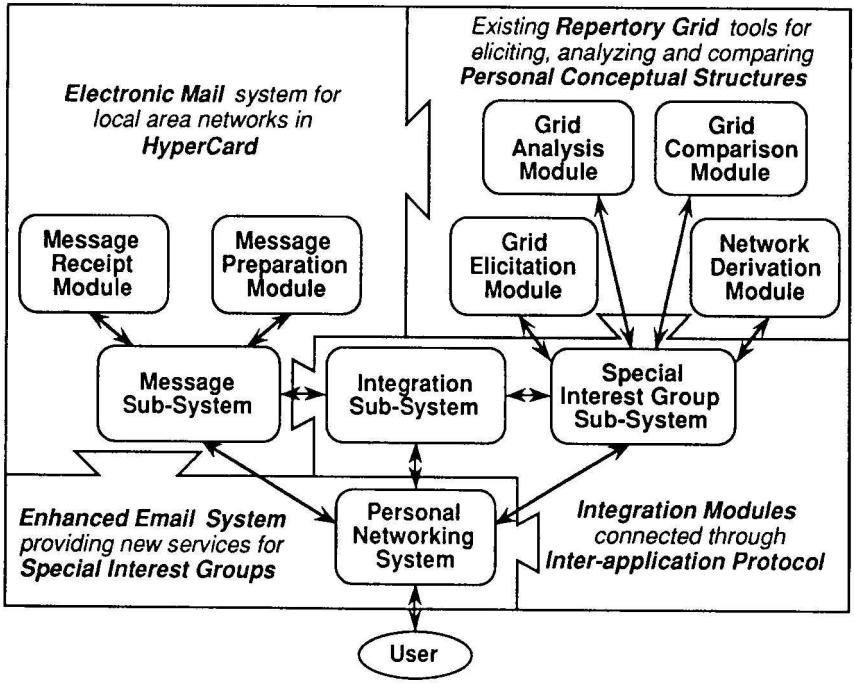


Fig.34 RepGrid–Net architecture

The current implementation is written for the Apple Macintosh and operates only on a local server network. Transferring the code to other personal computers and workstations involves the usual problems of portability of highly interactive user interfaces. Remote operation presents no major problems since the grids are compact data structures similar in size to electronic mail messages, and work is already under way to allow the system to operate over global networks. Thus, it is feasible to bring the systems described in this paper out of the laboratory and make them an integral part of communication networks where the medium begins to provide explicit support for structured discourse.

The conceptual structure graphic editor shown in Figure 6 is part of our ongoing development of an integrated knowledge support system architecture (GAINES 1987, 1991d, GAINES; SHAW 1987, GAINES; LINSTER 1990, GAINES; RAPPAPORT; SHAW 1989) of which RepGrid is one issued version, and RepGrid–Net another. Much of our work in recent years has gone into recoding our earlier systems as a class library in C++ to provide a strong software engineering base for increasingly complex experimental systems. We have adopted a *server* architecture also which is highly modular and allows specialist subsystems to be prototyped very rapidly, such as RepGrid–Net, GroupWriter,

a group word–processor supporting collaborative writing (MALCOLM; GAINES 1991), and Induct, a system for extracting conceptual models from large databases (GAINES 1989b).

CONCLUSIONS

The study of discourse within specialized communities with a common disciplinary, or goal–directed, focus requires methodologies for the empirical derivation and comparative analysis of both the conceptual and terminological structures used by participants. This paper has reported on research studies over the past sixteen years in which we have developed a theoretical framework, quantitative empirical methodologies, and computer–based analytical tools, for modeling terminological and conceptual relations in language use across communities.

Recent developments have been illustrated of on–line, computer–based tools for the interactive elicitation of concepts and terminology using graphic interfaces on personal computers. Some new extensions allowing the use of these tools over networks to support discourse and knowledge processes in dispersed communities have been described.

This paper has emphasized the close links between the theoretical foundations, linguistic and knowledge representation concepts, empirical methodology, its implementation in computer–based tools, and practical applications. It has illustrated the powerful tools now available for the study of quantitative linguistic phenomena in specialized communities.

It is a long, slow, and sometimes very painful, uphill climb, but we believe progress is being made towards our vision of an integrated family of computer–based tools, based on well–founded theories and methodologies, that support human knowledge processes. We envision this support both in the laboratory, experimental sense of increasing the depth and precision of data collection and analysis tools, and in the practical sense of extending human capabilities.

ACKNOWLEDGEMENTS

Financial assistance for this work has been made available by the Natural Sciences and Engineering Research Council of Canada. The RepGrid system was made available by the Centre for Person Computer Studies. We are grateful to Burghard Rieger for this opportunity to report to colleagues in the linguistic community the current state of these research studies.

REFERENCES

Adams–Webber, J.R. (1979): *Personal Construct Theory*. Chichester: Wiley.

Anglin, J.M. (1977): *Word, Object, and Conceptual Development.* New York: Norton.

Boose, J.H. (1984): Personal construct theory and the transfer of human expertise. In: *Proceedings AAAI-84*. California: American Association for Artificial Intelligence, 27-33.

Boose, J.H.; Bradshaw, J.M. (1987): Expertise transfer and complex problems: using AQUINAS as a knowledge acquisition workbench for knowledge-based systems. In: *International Journal of Man–Machine Studies* 26, 3-28.

Boose, J.H.; Gaines, B.R. (Eds.) (1988): *Knowledge Acquisition Tools for Expert Systems*. London: Academic Press.

Brown, G.S. (1969): *Laws of Form*. London: George Allen & Unwin.

Clancey, W.J. (1990): The frame of reference problem in the design of intelligent machines. In: K. van Lehn; A. Newell: *Architectures for Intelligence: The Twenty-Second Carnegie Symposium on Cognition.* Hillsdale: LEA.

CPCS (1990): *RepGrid Manual.* Centre for Person–Computer Studies, 3019 Underhill Dr NW, Calgary, Alberta, Canada T2N 4E4.

Cruse, D.A. (1986): *Lexical Semantics*. Cambridge: University Press.

Gaines, B.R. (1987): An overview of knowledge acquisition and transfer. In: *International Journal of Man–Machine Studies* 26(4) (April), 453-472.

Gaines, B.R. (1988): Knowledge acquisition systems for rapid prototyping of expert systems. In: *INFOR* 26(4) (November), 256-285.

Gaines, B.R. (1989a): Social and cognitive processes in knowledge acquisition. In: *Knowledge Acquisition* 1(1) (March), 251-280.

Gaines, B.R. (1989b): An Ounce of Knowledge is Worth a Ton of Data: Quantitative Studies of the Trade–Off between Expertise and Data based on Statistically Well–Founded Empirical Induction. In: *Proceedings of 6th International Workshop on Machine Learning*. Ithaca: Cornell University.

Gaines, B.R. (1991a): Between neuron, culture and logic: explicating the cognitive nexus. In: *Proceedings of ICO'91: Cognitive Science: Tools for the Development of Organizations*. Montreal: GIRICO, 222-32.

Gaines, B.R. (1991b): An interactive visual language for term subsumption visual languages. In: *IJCAI'91: Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*. San Mateo, California: Morgan Kaufmann.

Gaines, B.R. (1991c): Empirical investigation of knowledge representation servers: design issues and applications experience with KRS. In: *AAAI Spring Symposium: Implemented Knowledge Representation and Reasoning Systems*. Stanford (March), 87-101 — also *SIGART Bulletin* 2(3), 45-56.

Gaines, B.R. (1991d): Organizational modeling and problem solving using an object–oriented knowledge representation server and visual language. In: *COCS'91: Proceedings of Conference on Organizational Computing Systems*. ACM Press.

Gaines, B.R.; Linster, M. (1990): Integrating a knowledge acquisition tool, an expert system shell and a hypermedia system. In: *International Journal of Expert Systems Research and Applications* 3(2), 105-129.

Gaines, B.R., Rappaport, A.; Shaw, M.L.G. (1989): A heterogeneous knowledge support system. In: Boose, J.H.; Gaines, B.R. (Eds.): *Proceedings of the Fourth AAAI Knowledge Acquisition for Knowledge–Based Systems Workshop*. Banff (October), 13-1 - 13-20.

Gaines, B.R.; Shaw, M.L.G. (1980): New directions in the analysis and interactive elicitation of personal construct systems. In: *International Journal of Man–Machine Studies* 13, 81-116.

Gaines, B.R.; Shaw, M.L.G. (1984): Hierarchies of distinctions as generators of system theories. In: Smith, A.W. (Ed.): *Proceedings of the Society for General Systems Research International Conference*. California: Intersystems, 559-566.

Gaines, B.R.; Shaw, M.L.G. (1987): Knowledge support systems. In: *ACM MCC–University Research Symposium*. Austin, Texas: MCC, 47-66.

Gaines, B.R.; Shaw, M.L.G. (1989): Comparing the conceptual systems of experts. In: *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*. San Mateo, California: Morgan Kaufmann, 633-638.

Gaines, B.R.; Shaw, M.L.G. (1990): Cognitive and logical foundations of knowledge acquisition. In: Boose, J.H.; Gaines, B.R. (Eds.): *Proceedings of the Fifth AAAI Knowledge Acquisition for Knowledge–Based Systems Workshop*. Banff (November), 9-1 - 9-25.

Gomez, F. (1990): Knowledge acquisition from natural language for expert systems based on classification problem-solving methods. In: *Knowledge Acquisition* 2(2), to appear.

Habermas, J. (1981): *The Theory of Communicative Action: Reason and the Rationalization of Society*. Boston: Beacon Press.

Hall, E.T. (1959): *The Silent Language*. New York: Doubleday.

Hatcher, W.S. (1982): *The Logical Foundations of Mathematics*. Oxford: Pergamon Press.

Hausser, R. (1987): *Computation of Language*. Berlin: Springer.

Heller, A. (1990): *Can Modernity Survive?* Berkeley: University of California Press.

Kelly, G.A. (1955): *The Psychology of Personal Constructs*. New York: Norton.

Kerr, E.B.; Hiltz, S.R. (1982): *Computer-Mediated Communication Systems*. New York: Academic Press.

Lloyd, G.E.R. (1966): *Polarity and Analogy: Two Types of Argumentation in Early Greek Thought*. Cambridge: University Press.

Malcolm, N.; Gaines, B.R. (1991): A minimalist approach to the development of a word processor supporting group writing activities. In: *COCS'91: Proceedings of Conference on Organizational Computing Systems*. ACM Press.

Mancuso, J.C.; Shaw, M.L.G. (Eds.) (1988): Cognition and Personal Structure: Computer Access and Analysis. New York: Praeger Press.

Marcus, R.B. (1962): Interpreting quantification. In: *Inquiry* 5, 252-259.

Maybury-Lewis, D.; Almagor, U. (Eds.) (1989): The Attraction of Opposites: Thought and Society in the Dualistic Mode. Ann Arbor: University of Michigan Press.

Nosek, J.T.; Roth, I. (1990): A comparison of formal knowledge representations as communication tools: predicate logic vs semantic network. In: *International Journal of Man–Machine Studies* 33, 227-239.

Pask, G. (1975): *Conversation, Cognition and Learning*. Amsterdam: Elsevier.

Popper, K.R. (1968): Epistemology without a knowing subject. In: Van Rootselaar, B. (Ed.): *Logic, Methodology and Philosophy of Science III*. Amsterdam: North–Holland, 333-373.

Popper, K.R. (1974): Autobiography of Karl Popper. In: Schilpp, P.A. (Ed.): *The Philosophy of Karl Popper*. La Salle, Illinois: Open Court, 3-181.

Quartermain, J. S. (1990): *The Matrix: Computer Networks and Conferencing Systems Worldwide*. Digital Press.

Quillian, M.R. (1968): Semantic memory. In: Minsky, M. (Ed.): *Semantic Information Processing*. Cambridge, Massachusetts: MIT Press, 216-270.

Rieger, B.B. (1981): *Empirical Semantics*. Bochum: Brockmeyer Press.

Rieger, B.B. (1991): On distributed representation in word semantics. TR-91-012, Dept of Computational Linguistics, University of Trier.

Shaw, M.L.G. (1980): *On Becoming a Personal Scientist.* London: Academic Press.

Shaw, M.L.G. (Ed.) (1981): *Recent Advances in Personal Construct Technology.* London: Academic Press.

Shaw, M.L.G. (1985): Communities of knowledge. In: Epting, F.; Landfield, A.W. (Eds.): *Anticipating Personal Construct Psychology.* Lincoln: University of Nebraska Press, 25-35.

Shaw, M.L.G. (1988): An interactive knowledge–based system for group problem-solving. In: *IEEE System, Man and Cybernetics* 18, 610-617.

Shaw, M.L.G.; Gaines, B.R. (1981): Exploring personal semantic space. In: Rieger (1981).

Shaw, M.L.G.; Gaines, B.R. (1983): A computer aid to knowledge engineering. In: *Proceedings of British Computer Society Conference on Expert Systems.* Cambridge (December), 263-271.

Shaw, M.L.G.; Gaines, B.R. (1987): KITTEN: Knowledge Initiation & Transfer Tools for Experts & Novices. In: *International Journal of Man–Machine Studies* 27(3), 251-280.

Shaw, M.L.G.; Gaines, B.R. (1989): A methodology for recognizing conflict, correspondence, consensus and contrast in a knowledge acquisition system. In: *Knowledge Acquisition* 1(4), 341-363.

Shaw, M.L.G.; Gaines, B.R. (1991a): Supporting personal networking through computer networking. In: *Proceedings of CHI'91: Human Factors in Computing Systems.* New York: ACM Publications, 437-438.

Shaw, M.L.G.; Gaines, B.R. (1991b): Extending electronic mail with conceptual modeling to provide group decision support. In: *COCS'91: Proceedings of Conference on Organizational Computing Systems.* ACM Press (November).

Shaw, M.L.G.; Woodward, J.B. (1987): Validation in a knowledge support system: construing and consistency with multiple experts. In: *International Journal of Man–Machine Studies* 29, 329-350.

Simmel, G. (1950): *The Sociology of Georg Simmel* (trans. Wolff, K.H.). New York: Free Press.

Wolff, K.H. (1976): *Surrender and Catch: Experience and Enquiry Today.* Dordrecht: D.Reidel.

Zadeh, L.A. (1964): The concept of state in system theory. In: Mesarovic, M.D. (Ed.): *Views on General Systems Theory: Proceedings of the Second Systems Symposium at Case Institute of Technology*. New York: John Wiley.