Computing With Words Is an Implementable Paradigm: Fuzzy Queries, Linguistic Data Summaries, and Natural-Language Generation

Janusz Kacprzyk, *Fellow, IEEE*, and Sławomir Zadrożny

***Abstract*—We point out some relevant issues that are related to the computing-with-words (CWW) paradigm and argue for an urgent need for a new, nontraditional look at the area, since the traditional approach has resulted in very valuable theoretical research results. However, there is no proper exposure and recognition in other areas to which CWW belongs and can really contribute, notably natural-language processing (NLP), in general, and natural-language understanding (NLU) and natural-language generation (NLG), in particular. First, we present crucial elements of CWW, in particular Zadeh's protoforms, and indicate their power and stress a need to develop new tools to handle more modalities. We argue that CWW also has a high implementation potential and present our approach to linguistic data(base) summaries, which is a very intuitive and human-consistent natural-language-based knowledge-discovery tool. Special emphasis is on the use of Zadeh's protoform (prototypical form) as a general form of a linguistic data summary. We present an extension of our interactive approach, which is based on fuzzy logic and fuzzy database queries, to implement such linguistic summaries. In the main part of the paper, we discuss a close relation between linguistic summarization in the sense considered and some basic ideas and solutions in NLG, thus analyzing possible common elements and an opportunity to use developed tools, as well as some inherent differences and difficulties. Notably, we indicate a close relation of linguistic summaries that are considered to be some type of an extended template-based, and even a simple phrase-based, NLG system and emphasize a possibility to use software that is available in these areas. An important conclusion is also an urgent need to develop new protoforms, thus going beyond the classical ones of Zadeh. For illustration, we present an implementation for a sales database in a computer retailer, thereby showing the power of linguistic summaries, as well as an urgent need for new types of protoforms. Although we use linguistic summaries throughout, our discussion is also valid for CWW in general. We hope that this paper—which presents our personal view and perspective that result from our long-time involvement in both theoretical work in broadly perceived CWW and real-world implementations—will trigger a discussion and research efforts to help find a way out of a strange situation in which, on one hand, one can clearly see that CWW is related to words (language) and computing and, hence, should be part of broadly perceived mainstream computational linguistics, which lack tools to handle imprecision. These tools can be provided by CWW. Yet, CWW is practically unknown to these communities and is not mentioned or cited, and—reciprocally—even the top people in CWW do not refer to the results that are obtained in these areas. We hope that our paper, for the benefit of both the areas, will help bridge this gap that results from a wrong and dangerous fragmentation of science.**

***Index Terms*—Computational linguistics, computing with words (CWW), data mining, fuzzy logic, fuzzy query, linguistic summarization, natural-language generation (NLG), protoform.**

I. Introduction

THE PARADIGM of computing with words (CWW), which was introduced by Zadeh in the mid-1990s, had attracted much attention, notably in fuzzy logic and soft computing. Many powerful results were obtained, and some existing models and applications were redefined and presented in the perspective of CWW. An example is database querying (cf., [1] and [2]) or linguistic data summarization [3]. Moreover, some conceptually new approaches to CWW were proposed, notably by Mendel [4]–[6] and Mendel and Wu [7], who used the type-2 fuzzy sets and provided a novel conceptual perspective, as well as by Türksen [8], [9]; see also, e.g., [10]–[12] for more general considerations. Very relevant for CWW are also works by Herrera and his collaborators, for instance, Herrera and Martínez [13], [14] and Herrera and Herrera-Viedma [15], [16], to name a few. Different perspectives of CWW can be found, for instance, in [17]–[24], but they do not follow our line of reasoning.

The edited volumes by Zadeh and Kacprzyk [25] are a full coverage of foundations, theoretical developments, and applications of CWW. Since that time, the paradigm of CWW has attracted much attention and has been a subject of intensive research. We will use here the term CWW, and not "computing with words and perceptions (CWP)," because although CWW is closely related to perceptions, we do not want to get involved with a more fundamental discussion on perceptions dealt with in psychology, cognitive science, neuroscience, etc. However, the concept of perceptual computer/computing by Mendel [4], Wu and Mendel [26], etc., may provide, in this respect, a new insight.

Unfortunately, in spite of inherent potentials of CWW, it has been practically unnoticed by large scientific communities that are involved in computational-linguistics-related areas, notably natural-language processing (NLP) in general and natural-language understanding (NLU) and natural-language generation (NLG) in particular; these communities have produced theoretic results, algorithms, software, and real-world applications. This can be viewed strange because CWW could offer effective and efficient tools and techniques to handle

Manuscript received May 9, 2009; revised August 31, 2009; accepted November 24, 2009. Date of publication January 15, 2010; date of current version May 25, 2010.

The authors are with the Systems Research Institute, Polish Academy of Sciences, 01-447 Warsaw, Poland (e-mail: kacprzyk@ibspan.waw.pl; zadrozny@ibspan.waw.pl).

Digital Object Identifier 10.1109/TFUZZ.2010.2040480

imprecision of meaning that is crucial in natural language yet not managed by traditional approaches. Without the acceptance of CWW as a proper paradigm for dealing with (aspects of) natural language by the relevant research communities, the real development and exposure of CWW can be, in our opinion, difficult.

A way out of this difficult situation should be sought, and some reflections and solutions are proposed here. First, virtually all works that are related to CWW have, so far, emphasized theoretical aspects, which are mostly related to (fuzzy) logic and approximate reasoning (cf., [10]). Much less attention has been put, first, on implementations, and, second, on a relation to linguistics, like language representation, processing, and generation. Even the applied works used CWW "as is," i.e., with no account for relations to computational-linguistics-related areas. As an example, the work of Kacprzyk and Zadrożny [1], [27]–[36] can be cited in which fuzzy querying, linguistic data summaries, or even novel concepts in group decision making and consensus formation were presented in the perspective of CWW, thus advocating many advantages, mainly due to a more uniform and general line of reasoning and analysis, in particular using Zadeh's protoforms. One of the first attempts to go beyond this traditional way of viewing CWW as a self-contained field was that of Kacprzyk and Zadrożny [27], who proposed a preliminary suggestion of a close link of some relevant developments in CWW with a relevant and rapidly developing area of NLG. The purpose of this paper is to further elaborate on this claim by the authors and propose a new way of looking at the CWW paradigm as a useful, implementable paradigm to model and solve many problems in which natural language is an appropriate way to express values, relations, inference schemes, chains of reasoning, etc. We claim that there will be no success in the advancement and proliferation of CWW without taking into account developments in other areas that involve similar natural-language-related aspects, notably NLP. We think that these areas need proper tools to represent and handle imprecision that is inherent in natural language, and these tools can indeed be provided by CWW. This will be a huge research undertaking if we take into account what has already been accomplished in NLP, and in this paper, we cannot elaborate on all these aspects. We only wish to trigger a discussion and change the present mindset, and we will specifically deal with NLG only to show what can and should be done.

To accomplish our goal that, as we believe, would result in some rethinking and reorientation of a research attitude and paradigm in the community that works on CWW, and—as a result—imply a wider acceptance of CWW in the large computational-linguistics-related communities, we will try to concentrate on some specific problems in which CWW provides new vistas and shows its power. Then, we will provide some insight and a new look into how some relevant elements of CWW can be recast in the context of NLG. All this will be based on our works on linguistic data summaries and fuzzy querying, which will first be surveyed to provide a point of departure for the main, conceptually new part of this paper, i.e., a suggestion on how to redefine, reorient, and reimplement these works in terms of what has already been done in NLG.

In Section II, we briefly present CWW from our point of view. Section III is concerned with a survey of our works on linguistic data summaries to show the power of CWW and fuzzy querying that can provide tools to implement linguistic summaries. In Section IV, which is the authors' original proposal, some better known approaches and tools that are employed in NLG are presented, and their relations to elements of CWW are indicated, with remarks on the existence of software that can be used. In Section V, our implementation of linguistic summaries is reconsidered in terms of what could be gained if CWW were recast to use elements of NLG.

II. Computing With Words

The paradigm of CWW, which was introduced by Zadeh in the mid-1990s [25], [37], [38], is a logical consequence of the following line of reasoning. Numbers are used in all traditional "precise" models of systems, reasoning schemes, algorithms, etc., but for the humans, the only fully natural means of articulation and communication is natural language. Therefore, maybe we could develop models, tools, techniques, algorithms, etc., that could operate on natural language (words) and can serve the same (or similar) purpose as their numerical counterparts, i.e., maybe instead of traditional computing with numbers (from measurements), it would be better to compute with words (from perceptions). Therefore, we may skip an "artificial'' interface and try to operate on what is human-specific: natural language; this is in line with the famous human-centric paradigm pioneered at the Massachusetts Institute of Technology by Dertouzos [39].

The very essence of CWW is related to natural language and its processing, which are the areas that have been developing successfully for years without the use of fuzzy logic or possibility theory, although it was obvious that natural language is imprecise *per se*, and hence, conventional crisp and formal tools could have difficulties. Moreover, CWW is a natural consequence of fuzzy logic and Zadeh's earlier works on a linguistic approach to the analysis of systems, the representation of hedges, PRUF, test score semantics, etc. (cf., [40] and [41]). Concerning fuzzy logic, (natural) language is practically the only vehicle to illustrate the semantics of its main concepts and operations. Therefore, even if we strongly base CWW on fuzzy logic, its basic philosophy and idea are much more general, although fuzzy logic provides perhaps the most effective and efficient tools and techniques; our opinion differs from Zadeh's [37] perspective in which fuzzy logic is practically a *sine qua non* condition for CWW. Basically, CWW is a necessity when the available information is too imprecise to use numbers and when there is a tolerance for imprecision to attain tractability, robustness, lower solution cost, or a better rapport with reality.

The point of departure in Zadeh's [37] CWW is the concept of a granule, which, in the fuzzy setting, is a fuzzy set of points that form a collection of elements that are drawn together by similarity. A word w is a label of a granule g, and g is the denotation of w. A granule g, which is the denotation of w, is viewed as a fuzzy constraint on a variable value. Particularly important in

CWW is the fuzzy-constraint propagation from premise to conclusion. Information is obviously conveyed by constraining the values of variables, and information consists of a collection of propositions that are expressed in some (quasi)natural language.

The essence of CWW is as follows: We have a collection of propositions that are expressed in a natural language, and, from this set, we wish to infer an answer to a query that is expressed in a natural language. The answer is also expressed in a natural language.

For the implementation of CWW, first, we need tools for a formal representation of linguistic terms, relations, etc., and this is provided by a precisiated natural language (PNL) [42]. A generalized constraint, which represents the meaning of a proposition p in a natural language, is written as

$$X \text{ isr } R$$

where X is a constrained variable that, in general, is implicit in p, R is the constraining relation that, in general, is implicit in p, and r is an indexing variable whose value identifies how R constrains X.

The variable r may have a different meaning, and the main types of generalized constraints are the following:

1) e: equal;
2) d: disjunctive (possibilistic);
3) c: conjunctive;
4) p: probabilistic;
5) λ: probability value;
6) u: usuality;
7) *rs*: random set;
8) *rsf*: random fuzzy set;
9) *fg*: fuzzy graph;
10) *ps*: rough set, etc.

For instance, when r takes the value c, the constraint is conjunctive. In this case, "X isc R" means that if the grade of membership of u in R is $\mu(u)$, then $X = u$ has truth value $\mu(u)$.

The constraint "X isu R," i.e., "usually(X is R)," means that "Prob$\{X$ is $R\}$ is usually." Here, "X is R" is a fuzzy event and "usually" is its fuzzy probability, i.e., the possibility distribution of its crisp probability. We will come back later to the usuality constraints as they are extremely important in many applications, notably for our discussion.

The role of "r" is to add a fine shade of meaning to a basic constraint. This is the essence of modality in natural language (cf., [43]), which allows us to express, for instance, degrees of

1) usuality—how frequently something occurs or is true;
2) probability, possibility, or certainty—the likelihood of something happening or being the case;
3) obligation or necessity—how necessary it is for things to be done or to be in a certain way;
4) ability—the ability of someone or something to do something;
5) inclination—the inclination or willingness of someone to do something, etc.

Note that this is much richer than what CWW can provide so far, but modalities like obligation, ability, or inclination are very difficult to handle, and in our opinion, results from computational linguistics should be more explicitly employed.

The usuality constraint is crucial because, in most cases, we seek some "regularities," "normal/typical" relations, i.e., those that usually happen (occur). However, one can notice that the semantics of usuality calls for the use of linguistic quantifiers like "most," "almost all," etc. Fuzzy-logic-based calculi of linguistically quantified propositions are intuitively appealing and computationally efficient. We will later concentrate on the specific case where the usuality plays a crucial role and how it implies an extremely powerful concept of a linguistic data summary, which is perhaps the most practically relevant concept in CWW.

To summarize, on one hand, CWW may be viewed as powerful with respect to a natural and simple representation of inherently imprecise elements in natural language, and by virtue of this, it should be found useful by the computational linguistics community. On the other hand, CWW still does not provide a means to handle all relevant aspects of natural language, as with some important modalities, and this can hopefully be overcome by following the path that is advocated in this paper: to more explicitly use what has been done in computational-linguistics-related areas.

III. Linguistic Data Summaries

We will now briefly present our works on the linguistic summaries of (numerical) data: one of the key areas in which CWW proved to be operational, useful, and implementable. We will point out some crucial elements of CWW that are employed therein and indicate difficulties in their handling; this will provide a point of departure and rationale for the next section in which we will advocate that the use of NLG tools and techniques may greatly help in this respect.

A. Introductory Remarks

Data summarization is one of the basic capabilities that is needed by any "intelligent" system meant to operate in real life with an abundance of data that is beyond human cognition and comprehension. Since, for a human being, the only fully natural means of communication is natural language, a linguistic (by some sentence(s) in a natural language) data summarization would be very desirable and human-consistent. For instance, having a personnel database, a statement (linguistic summary) like "almost all younger and well-qualified employees are well paid" would be useful and human-consistent. Unfortunately, data summarization is still, in general, an unsolved problem, despite vast research efforts.

We will show the use of linguistic data(base) summaries that were introduced by Yager [44]–[47], advanced by Kacprzyk and Yager [48], Kacprzyk *et al.* [49], and Zadrożny and Kacprzyk [50], and implemented by Kacprzyk and Zadrożny [2], [29]–[34], [51]–[53]. We will derive linguistic data summaries as linguistically quantified propositions as, e.g., "most of the employees are young and well paid," with a degree of validity (truth); some other validity (quality) indicators can also be added (cf., [54], or [48] and [49]). For other approaches to the linguistic summarization of databases, cf., [55]–[58]. One should also mention linguistic data summaries using type-2 fuzzy sets, as proposed by Niewiadomski [59]. Very relevant results may also be obtained using some other related techniques as, e.g., those of fuzzy association rules [60], [61], which are recast in the context of linguistic summaries by Kacprzyk and Zadrożny [34], [53].

We employ Kacprzyk and Zadrożny's [29], [33] idea of an interactive approach to linguistic summaries by using Kacprzyk and Zadrożny's [1], [2], [62], [63] FQUERY for Access, which is a fuzzy querying add-on (see also [64] for a comprehensive review of fuzzy querying). The line of reasoning is that since a fully automatic generation of linguistic summaries is not feasible at present, some user interaction is assumed to determine a class of summaries of interest, and this is done via this fuzzy querying add-on.

First, extending the work of Kacprzyk and Zadrożny [52], we show that by relating various types of linguistic summaries to fuzzy queries, with various known and sought elements, we can arrive at a hierarchy of Zadeh's [65] protoforms of linguistic data summaries. This indeed seems to be a very powerful conceptual idea.

In Section IV, we discuss an intrinsic relation between data summarization and NLG, which is a rapidly developing area in which strong results, algorithms, and software are available. This can open new vistas for the implementation of linguistic summaries and, through this, for the implementability of CWW. In particular, we comment in more detail on the use of Zadeh's protoforms because, on one hand, they provide a very useful and powerful general tool, but—on the other hand—they imply the use of some simpler and less-sophisticated architectures, solutions, and tools that are known in NLG, which is clearly both good and bad.

B. Linguistic Summaries Using Fuzzy Logic With Linguistic Quantifiers

In Yager's [44] approach, we have (we use here the source terminology) 1) V, which is quality (attribute) of interest, e.g., salary in a database of workers; 2) a set of objects (records) that manifest quality V, e.g., the set of workers (hence, $V(y_i)$ are the values of quality V for objects y_i); and 3) $D = \{V(y_1), \ldots, V(y_n)\}$ is a set of data (the "database" in question).

A linguistic summary of a dataset consists of

1) a summarizer S (e.g., young)
2) a quantity in agreement Q (e.g., most);
3) truth degree T, e.g., 0.7, as, e.g., "T (most of employees are young) $= 0.7$."

The summarizer S is a linguistic expression that is semantically represented by a fuzzy set. The meaning of S, i.e., its corresponding fuzzy set is, in practice, subjective and may be either predefined or elicited from the user (as shown later).

A simple one-attribute-related summarizer that is exemplified by "young" serves the purpose of introducing the concept of a linguistic summary but is of a lesser practical relevance. It can be extended for some confluence of attribute values as, e.g., "young and well paid," and then to more complicated combinations. The most interesting are nontrivial, human-consistent summarizers (concepts), e.g., productive workers, stimulating work environment, difficult orders, etc., which involve complicated combinations of attributes, e.g., a hierarchy (not all attributes are of the same importance), the attribute values are ANDed and/or ORed, k out of n, most of them should be accounted for, etc.

The quantity in agreement, i.e., Q, is a proposed indication of the extent to which the data satisfy the summary, and a linguistic term represented by a fuzzy set is employed. Basically, two types of a linguistic quantity can be used in an agreement: 1) absolute, e.g., "about 5," "several", and 2) relative, e.g., "a few," "most," almost all," etc., and they are equated with the so-called fuzzy linguistic quantifiers (cf., [66] and [67]) that can be handled by fuzzy logic.

The calculation of the truth degree is equivalent to the calculation of the truth value (from [0.1]) of a linguistically quantified statement that may be done using Zadeh's [66] calculus of linguistically quantified propositions (cf., [25]), and this will be used here; cf., Yager's [68] ordered weighted averaging (OWA) operators (cf., [69]).

A linguistically quantified proposition, like "most experts are convinced," is written as "*Qy*'s are F," where Q is a linguistic quantifier (e.g., most), $Y = \{y\}$ is a set of objects (e.g., experts), and F is a property (e.g., convinced). Importance B may be added, thereby yielding "*QBy*'s are F," e.g., "most (Q) of the important (B) experts (y's) are convinced (F)." The problem is to find truth(Qy's are F) or truth(QBy's are F), respectively, knowing that truth(y is F)$\forall y \in Y$, which is done using Zadeh's calculus of linguistically quantified propositions as follows.

Property F and importance B are fuzzy sets in Y, and a (proportional, nondecreasing) linguistic quantifier Q is a fuzzy set in [0, 1] as, e.g.,

$$\mu_Q(x) = \begin{cases} 1, & \text{for } x \geq 0.8 \\ 2x - 0.6, & \text{for } 0.3 < x < 0.8 \\ 0, & \text{for } x \leq 0.3. \end{cases} \tag{1}$$

Then, according to Zadeh [66]

$$\text{truth}(Qy\text{'s are } F) = \mu_Q\left[\frac{1}{n}\sum_{i=1}^{n}\mu_F(y_i)\right] \tag{2}$$

$$\text{truth}(QBy\text{'s are } F) = \mu_Q\left[\frac{\sum_{i=1}^{n}(\mu_B(y_i) \wedge \mu_F(y_i))}{\sum_{i=1}^{n}\mu_B(y_i)}\right] \tag{3}$$

where "$\wedge$" denotes the minimum operator; other operations, notably t-norms, can be used clearly.

Zadeh's calculus of linguistically quantified propositions makes it possible to formalize more complex linguistic summaries, which are exemplified by "most (Q) newly hired (K) employees are young (S)," i.e., a linguistic summary may also contain a *qualifier K*.

The basic validity criterion, i.e., the truth of (2) and (3), is certainly the most important, but it does not grasp all aspects of a linguistic summary. As to some other quality (validity) criteria, e.g., Yager [44] proposed a measure of informativeness, and, then, five additional measures have been proposed by Kacprzyk and Yager [48] and Kacprzyk *et al.* [49]: truth, degrees of imprecision, covering, appropriateness, and a length of a summary.

An obvious problem is how to generate the best summary (or summaries). An exhaustive search can obviously be computationally prohibitive. We deal with this question in the next section.

The earlier approach to linguistic data summaries was implemented to support decision making in a computer retailer (see [3, Sec. V]), summarization of Web server logs [70], and linguistic summarization of an investment (mutual) fund quotations [71]–[73].

C. Fuzzy Querying, Linguistic Summaries, and Their Protoforms

As a promising attempt to operationally generate the linguistic data summaries, Kacprzyk and Zadrożny [33] proposed interactivity, i.e., user assistance, in the definition of summarizers (indication of attributes and their combinations) via a user interface of a fuzzy querying add-on. The roots of this approach are our previous papers on the use of fuzzy logic in querying databases (cf., [74] and [75]) and the work of Kacprzyk *et al.* [76]) via imprecise requests. This has motivated the development of the whole family of fuzzy querying interfaces, notably our FQUERY for Access package. Thus, such a fuzzy querying interface, on its own, is already a perfect example of an implementation of the CWW paradigm.

FQUERY for Access is an add-in that makes possible the use of fuzzy linguistic terms in queries, notably 1) fuzzy values, which are exemplified by "low" in "profitability is low"; 2) fuzzy relations, which are exemplified by "much greater than" in "income is much greater than spending"; and 3) linguistic quantifiers, which are exemplified by "most" in "most conditions have to be met."

Linguistic quantifiers provide for a flexible aggregation of simple conditions. In FQUERY for Access, the fuzzy linguistic quantifiers are defined as in Zadeh's calculus of linguistically quantified propositions (cf., Section III), although the OWA operators are also possible (cf., [18]–[20], [28], and [56] for details).

Obviously, fuzzy queries directly correspond to summarizers in linguistic summaries. Thus, the derivation of a linguistic summary may proceed in an interactive (user-assisted) way as 1) the user formulates a set of linguistic summaries of interest (relevance) using the fuzzy querying add-on; 2) the system retrieves records from the database and calculates the validity of each summary that is adopted; and 3) a most appropriate linguistic summary is chosen.

The concept of a protoform in the sense of Zadeh [65], [77] is highly relevant in this context. First of all, a protoform is treated as an abstract prototype of a linguistic summary

$$\text{“}Q\ Y\text{'s are } S\text{”} \tag{4}$$

$$\text{“}Q\ KY\text{'s are } S\text{”} \tag{5}$$

where Y is a set of objects, K is a qualifier, and S is a summarizer.

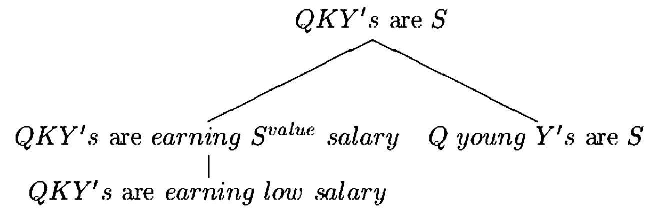


Fig. 1. Example of a part of a hierarchy of protoforms.

For the generation of linguistic summaries, it is convenient to consider the summarizer (and the qualifier) as an abstract fuzzy logic statement "X IS A," where X is a placeholder for an attribute of objects in Y, and A is a placeholder for a fuzzy set (linguistic term) that determines its value. Instantiated summarizer/qualifier can be, e.g., "age IS young," "salary IS low," and "salary IS A." Two former summarizers are fully instantiated, while the latter still contains an abstract form of the attribute value (A).

Evidently, as protoforms may form a hierarchy, we can define lower level (less abstract) protoforms, for instance, replacing Q by a specific linguistic quantifier "most," and we obtain

$$\text{“Most } Y\text{'s are } S\text{”} \tag{6}$$

$$\text{“Most } KY\text{'s are } S\text{”.} \tag{7}$$

Thus, we propose to use Zadeh's protoforms as a fundamental element of the user interface, namely, the user picks up a protoform of a linguistic summary from the earlier hierarchy, and then, the system instantiates the selected protoform in all possible ways, thus replacing abstract symbols (if any) that denote its elements with concrete fuzzy values and linguistic quantifiers that are stored in a dictionary of linguistic terms maintained by the fuzzy-querying module. An example of a part of such a hierarchy of protoforms is shown in Fig. 1. At the top of the hierarchy, we have a completely abstract protoform that corresponds to (5); in a protoform to the right, the qualifier K is instantiated to "age IS young"; in a left branch of the hierarchy, summarizer S is first instantiated to "salary IS S^{value}," i.e., the attribute of the summarizer is selected to be "salary," but its value is not determined; then, this protoform is further instantiated so that the summarizer is fully specified using "low" as the value of "salary." Thus, if the user picks up the top protoform, the system has to instantiate it using all possible combinations of the linguistic quantifiers, attributes for the summarizer and qualifier, and their corresponding fuzzy values. However, if the bottom left protoform is chosen, then only the combinations of linguistic quantifier and qualifier have to be checked, and the summarizer will be fixed.

Basically, the more abstract forms of protoforms correspond to cases in which we assume less about the summaries sought. There are two limit cases where we assume: 1) a totally abstract top protoform or 2) that all elements of a protoform are given on the lowest level of abstraction, i.e., all attributes and all linguistic terms that express their values are fixed. In case 1, data

TABLE I
CLASSIFICATION OF LINGUISTIC SUMMARIES

Type	Given	Sought	Remarks
1	S	Q	Simple summaries through ad-hoc queries
2	S B	Q	Conditional summaries through ad-hoc queries
3	Q $S^{structure}$	S^{value}	Simple value oriented summaries
4	Q $S^{structure}$ B	S^{value}	Conditional value oriented summaries
5	Nothing	S B Q	General fuzzy rules

summarization using a naïve full-search algorithm would be extremely time-consuming but might produce an interesting, unexpected view on data. In case 2, the user is, in fact, guessing a good candidate summary, but the evaluation is fairly simple, which is equivalent to the answering of a (fuzzy) query. Thus, case 2 refers to *ad hoc* queries. This may be shown as in Table I, where five basic types of linguistic summaries are provided, which correspond to protoforms of an increasingly abstract form.

In Table I, $S^{\text{structure}}$ denotes that attributes and their connection in a summary are known, while S^{value} denotes the values of the attributes sought.

Type-1 summaries may be obtained easily by a simple extension of fuzzy querying. The user has to construct a query and a candidate summary, and it has to be determined what the fraction of rows that matches this query is and what linguistic quantifier best denotes this fraction. A type-2 summary is a straightforward extension of type 1. Type-3 summaries require much more effort as their primary goal is to determine typical or exceptional, depending on the quantifier, values of an attribute. A type-4 summary is meant to find typical (exceptional) values for some, possibly fuzzy, subset of rows. From the computational point of view, type-5 summaries represent the most general form considered here: fuzzy rules that describe dependencies between specific values of particular attributes. The summaries of types 1 and 3 have been implemented as an extension to Kacprzyk and Zadrożny's [29] FQUERY for Access. Two approaches to type-5 summaries generation have been proposed. First, a subset of such summaries may be obtained by analogy with the association rules concept and by employing their efficient algorithms. Second, genetic algorithms may be used to search the summaries' space.

Zadeh's protoforms are, therefore, a powerful conceptual tool because we can formulate many different types of linguistic summaries in a uniform way and devise a uniform and universal way to handle different linguistic summaries. However, from the point of view of the very purpose of this paper, there occurs an immediate question, namely, it is easy to see that such generality and uniformity are somehow contradictive to the richness and variety of natural language. This important issue is discussed in the next section, where we indicate that although Zadeh's protoforms are extremely powerful, they somehow set an unnecessary limit on the sophistication of possible NLG architectures and tools. This is even true for hierarchies of protoforms, and most probably, further research is needed to devise "metaprotoforms," which we advocate in the next section.

IV. RELATIONS TO NATURAL-LANGUAGE GENERATION: A NEW BREAKTHROUGH?

The linguistic data summaries—maybe the most intuitively appealing and the most promising instance of a real practical strength of CWW—have considerable expressive power as they can capture the very meaning of a set of data, can be derived quite effectively and efficiently, and have been implemented (as presented in the next section). As always, one can ask a natural question: Are they similar/related to what is done in other areas (sometimes much bigger and better developed) and, if so, can we use tools and techniques (software?) obtained in these areas?

The linguistic database summaries may immediately be viewed as having a strong resemblance to the concepts relevant to NLG, but for unknown reasons, this was never considered explicitly in more detail. The first indication of a relation to NLG was given by Kacprzyk *et al.* [72], where a reference to an NLG-based approach to the linguistic summarization of time series, i.e., the SumTime project at the University of Aberdeen, U.K. [78], [79], was made. Later, in 2009, Kacprzyk and Zadrożny [27] more explicitly suggested a relation to NLG. We will further elaborate on this problem here, indicate some important issues that are common to linguistic summarization and NLG, and propose some possible new research directions. We hope that this will trigger some research effort in both the fuzzy logic and NLG community that should result in valuable results, and—what may be even more important—in qualitatively new results, perhaps a synergistic collaboration, and implementations.

We will show that by relating various types of linguistic summaries to some templates of fuzzy queries, with various known (fixed) and sought (placeholder) elements, we can arrive at a hierarchy of Zadeh's [65], [77] protoforms of linguistic data summaries. We will discuss this and other similar concepts and solutions in linguistic summarization and NLG and indicate which elements of the former approach would be difficult to handle and implemented using the concepts and tools of the latter approach, and *vice versa*.

Basically, NLG is concerned with how one can automatically produce high-quality natural-language text from computer representations of information that is not in natural language. Therefore, we follow the "numbers-to-words" path here.

NLG may be viewed from many perspectives (cf., [80]), and in our context, it may be expedient to independently consider the tasks of generation and the process of generation. One can identify three types of tasks:

1) text planning;
2) sentence planning;
3) surface realization.

Text planners select the information to be included in the output and use it to form a proper text structure; sentence planners organize the content of sentences, notably ordering their parts, and surface realizers convert sentence-sized chunks of representation into grammatically correct sentences.

In our works on linguistic summarization, we have been, so far, mainly concerned with the text planning, since our approach is explicitly protoform-based. We literally consider a protoform of a linguistic summary to be fixed and specified (structurally), and the purpose of a protoform-based linguistic summarization is to determine appropriate linguistic values of the linguistic quantifier, qualifier, and summarizer. However, if the protoform is considered to be in a "metasense," i.e., when the summarizer concerns the linguistically defined values of a compound concept, like "productivity" in a personnel database, with productivity described by an ordered (e.g., through importance assignment) list of criteria, then we have to do some sort of sentence planning. The same concerns hierarchies of Zadeh's protoforms. Finally, the use of protoform-based linguistic summaries precludes the use of surface planning in the strict sense. This is an example of what was indicated in Section III that Zadeh's protoforms are very powerful and convenient in CWW but may be a limiting factor in many real-world applications as their structure is too restricted, notably as compared with the richness of natural language. A solution may clearly be to develop different kinds of protoforms, notably those that are not explicitly related to usuality but to other modalities. This is, however, not trivial. To subsume, the use of the sentence planning and surface realization would presumably produce more advanced linguistic summaries that could capture more fine shades of meaning, but it is not clear how these tasks can be accommodated within the simple and efficient, yet somewhat strict, Yager's concept of a linguistic summary and our heavily protoform-based approach.

Other authors use different classifications of generation tasks, but they do not essentially change the relations between CWW (specifically, linguistic summarization) and NLG. Due to the use of protoforms, almost all tasks are simple due to a predefined structure of summaries, but to make a full use of the power of NLG tools, which can do the aforementioned tasks, one should maybe again think about different types of protoforms. Another viable alternative is to use an interactive human–computer interface, as we have actually employed in our implementation of linguistic summaries. Generator processes can be classified along two dimensions: sophistication and expressive power, starting with inflexible canned methods and ending with maximally flexible feature combination methods. The canned text systems, which are used in many applications, notably simpler software systems that just print a string of words with no change (error messages, warnings, letters, etc.), are not interesting for us and maybe neither for CWW, in general.

Template-based systems that are used when a text (e.g., a message) must be produced several times with slight alterations are more sophisticated. The template approach is used mainly for multiple-sentence generation, particularly when texts are regular in structure, such as stock-market reports. In principle, our approach is similar in spirit to template-based systems. One can say that Zadeh's protoforms can be viewed as playing a role similar to templates. However, there is an enormous difference as the protoforms are much more general and may represent such a wide array of various "templates" that maybe it would be more proper to call them "metatemplates."

Phrase-based systems employ generalized templates. In such systems, a phrasal pattern is first selected to match the top level of the input, and then, each part of the pattern is expanded into a more specific phrasal pattern that matches some subpart of the input, etc., with the phrasal pattern replaced by one or more words. The phrase-based systems can be powerful and robust but are very difficult to build beyond a certain size, because of difficulties in a correct specification of phrasal interrelationships. It seems that our approach to linguistic summarization can be viewed, from some perspective, as a simple phrase-based system. It should also be noted that since protoforms may form hierarchies, we can imagine that both the phrase and its subphrases can be properly chosen protoforms. The calculi of linguistically quantified statements can be extended to handle such a hierarchic structure of phrases (statements) although, at the semantic level, the same difficulties remain as in the NLG approach, i.e., an inherent difficulty to grasp the essence of multisentence summaries with their interrelations. We think that Zadeh's protoforms, meant in a more general sense, for instance, as hierarchical protoforms or "metaprotoforms," can make the implementation of a phrase-based NLG system in the context of CWW viable. However, these more general types of protoforms have to be developed first.

Feature-based systems represent some extreme of the generalization of phrases. Each possible minimal alternative of expression is represented by a single feature. Generation proceeds by the incremental collection of features appropriate for each part of the input until the sentence is fully determined. The feature-based systems are the most sophisticated generators. Their idea is very simple as any distinction in language is defined as a feature, analyzed, and added to the system, but unfortunately, there is a tremendous difficulty in maintaining feature interrelationships and in the control of feature selection. It is not at all clear how the idea of feature-based systems and our approach to linguistic data summarization as well as CWW, in general, can be merged. However, if this were done, CWW would be able to deal with more elements of natural language, for instance, with an extended list of modalities.

To summarize, it seems that if we use NLG tools to implement linguistic summaries, one can either fairly easily build a single-purpose generator for any specific application or, with some difficulty, adapt an existing sentence generator to the application, with acceptable results. This would correspond to a simple protoform-based linguistic summary. However, we do not think that one could easily build a general-purpose sentence generator or a text planner, but first, some research on richer families of protoforms would be needed.

An extremely relevant issue, maybe a prerequisite for implementability, is domain-modeling. The main difficulty is that it is very difficult to link a generation system to a knowledge base or a database that is originally developed for some nonlinguistic purpose due to a possible considerable mismatch. The construction of appropriate taxonomies or ontologies can be of much help. So far, in our approach, domain knowledge is—at the conceptual level—in the specification of appropriate protoforms that are comprehensible or traditionally used (e.g., as structures of business reports) in a specific domain. Domain knowledge also plays an important role at the lower level of semantic interpretation (definition) of the linguistic terms. Each numerical attribute has to be associated with a range of values it takes. Then, depending on the context, the linguistic term will be interpreted differently, e.g., "low" in "low salary" and in "low temperature."

Formally, similar to this is the authors' approach [81] to the support of consensus reaching by using some ontologies that provide the conceptualization of both the process itself and the domain of the decision-making problem, maybe even augmented with some tools to the linguistic summarization of the dynamics of consensus reaching [73].

Generation choice criteria constitute a very important, yet difficult and still unsolved, problem. Unfortunately, natural language, with its almost infinite flexibility, demands far more from the input to a generator than can be provided so far. This problem will play an increasingly important role since generators will be able to say the same things in many ways, and some tools will have to be added to ensure that an appropriate text is produced. Various facilities of the NLG systems are needed here, which are exemplified by an ability to classify types of users, purposes of summaries, etc. Again, Zadeh's protoforms can play a significant role in reducing the number of possible forms of summaries saying the samethings and, hence, simplify the implementation.

Our approach to the linguistic summarization is not fully automatic as an interaction with the human being is employed. A natural desire would be to attain an even higher functionality of an automated summarization. This is a very difficult problem that involves many aspects. In an NLG module, we often distinguish two components. First, we determine what should be said, and this is done by a planning component that may produce an expression representing the content of the proposed utterance. On the basis of this representation, the syntactic generation component produces the actual output sentence(s).

Syntactic generation can be described as the problem of finding the corresponding string for an input logical form.

In our present version of linguistic summaries, these syntactic-generation problems do not play a significant role as the structure of our summaries, which are based on protoforms, is quite fixed and related to (meta)templates. However, templates may often be unavailable as we may not know exactly which types of protoforms are comprehensible in a given domain. Then, some syntactic generator might be a solution. We think that syntactic generation should play an important role in all kinds of linguistic summaries, and more generally, in CWW, in particular, in view of what we have said about the fact that the structural strictness of Zadeh's protoform may be a limiting factor, and more sophisticated and complex types of protoforms are urgently needed.

Another topic in NLG is the so-called deep generation. Basically, it usually involves, at a conceptual level, selecting the text content and imposing a linear order on this content. It may often be better to impose a looser structure on the text to be generated, and this gives rise to data-driven approaches, which are often based on some heuristics. In this direction, a still-unsolved problem is the selection of an appropriate granularity for the presentation of information. In our linguistic summaries, this problem has been solved in an *ad hoc* manner by using a structural granulation that is attained by the use of some specific protoforms and the granulation of linguistic values used throughout the summarization process by following the "golden rule" of using 7 ± 2 values (the Miller magic number) with very good results in practical applications.

To summarize this discussion of relations between our protoform-based approach to the linguistic data summarization and modern approaches, solutions, and tools available in NLG, we can state that, on one hand, we can find much inspiration from recent developments in NLG, notably in the adjustment of protoforms to a comprehensible and/or commonly used approach in a specific domain by employing some sentence and text-planning tools. Moreover, one can very clearly find a deep justification for the power of Zadeh's protoforms by showing their intrinsic relation to templates, or perhaps metatemplates, or even simple phrase-based systems.

On the other hand, it seems that NLG can benefit from the approach and solutions that are adopted in our approach by finding the conceptually and numerically operational means to grasp and handle the problem of imprecision of meaning that is so characteristic for natural language but has not been appropriately considered in NLG despite an urgent need.

Clearly, as the very purpose of our analysis was conceptual, we have attempted to give more general remarks and suggestions to advocate and trigger further research. From this point of view, since, in NLG, there is much commercial and open-source software available, the user can clearly see what can be done using a particular software package, and then, our indications and suggestions should help him/her in a practical use of that software to deal with linguistic summaries and, more generally, with CWW. This will be, for sure, decisive for the development of linguistic summarization based on CWW and fuzzy logic, which was often difficult due to a lack of the available software.

At the conceptual level, one can view the importance of our analysis of relations between the linguistic summaries and NLG as providing a proof for the very power of Zadeh's protoforms, even though it also clearly indicates an urgent need to go beyond the classic Zadeh's protoforms and develop new, more sophisticated, and less structurally rigid types of protoforms.

Clearly, the earlier conclusions have been drawn for linguistic summaries because we focused on them for clarity and illustrativeness and because they are one of the key examples of powerful and intuitively appealing applications of CWW and have been theoretically and algorithmically developed and implemented. However, these conclusions also seem to be valid, to a large extent, for CWW in general.

We discussed NLG because this is clearly related to linguistic summaries. However, this should be viewed from a wider perspective, namely, CWW should be considered in relation to various areas in NLP, notably to NLU and NLG. However, a comprehensive study along these lines is a huge research effort, and hence, we have decided to start this research by just indicating a strong relation of a specific, yet very relevant, area of CWW, i.e., linguistic summaries, to a specific, yet very relevant,

TABLE II
LINGUISTIC SUMMARIES THAT EXPRESS RELATIONS BETWEEN THE GROUP OF PRODUCTS AND COMMISSION

Summary
About ½ of sales of network elements is with a high commission
About ½ of sales of computers is with a medium commission
Much sales of accessories is with a high commission
Much sales of components is with a low commission
About ½ of sales of software is with a low commission
About ½ of sales of computers is with a low commission
A few sales of components is without commission
A few sales of computers is with a high commission
Very few sales of printers is with a high commission

TABLE III
LINGUISTIC SUMMARIES THAT EXPRESS RELATIONS BETWEEN THE FOLLOWING ATTRIBUTES: SIZE OF CUSTOMER, REGULARITY OF CUSTOMER (PURCHASING FREQUENCY), DATE OF SALE, TIME OF SALE, COMMISSION, GROUP OF PRODUCT, AND DAY OF SALE

Summary
Much sales on Saturday is about noon with a low commission
Much sales on Saturday is about noon for bigger customers
Much sales on Saturday is about noon
Much sales on Saturday is about noon for regular customers
A few sales for regular customers is with a low commission
A few sales for small customers is with a low commission
A few sales for one-time customers is with a low commission
Much sales for small customers is for non-regular customers

area of NLP, i.e., NLG. However, we hope that our first results are very promising and clearly indicate a new, prospective research direction.

V. IMPLEMENTATION

Now, we will briefly outline an implementation for a sales database of a small/medium computer retailer in southern Poland, which shows some best (in the sense of validity) linguistic summaries of interesting relations between attributes that concern sales transactions.

First, some relations between the commission and the type of goods sold are shown in Table II. The results can be very helpful for, e.g., negotiating commissions for various products sold.

Next, in Table III, some of the linguistic summaries that express relations between the number of customers, regularity that customers make a purchase, date of sale, time of sale, commission, group of product, and day of sale are shown. This is an example of the most sophisticated form of linguistic summaries that are accommodated in the system briefly described in Section III.

TABLE IV
LINGUISTIC SUMMARIES THAT EXPRESS RELATIONS BETWEEN GROUP OF PRODUCTS, TIME OF SALE, TEMPERATURE, PRECIPITACION, AND TYPE OF CUSTOMERS

Summary
Very few sales of software is in hot days to individual customers
About 1/2 of sales of accessories is in rainy days on weekends by the end of the year
About 1/3 of sales of computers is in rainy days to individual customers

The system does not present just one linguistic summary but a set of them: not too many and not too few and leave their interpretation, usefulness, etc., to the user. This is—in our opinion—very important as it may guarantee user's autonomy that is relevant in decision support.

The summaries that are shown concern data from the company's own (internal) database, but the company operates in an environment (economic, climatic, social, etc.). The inclusion of weather data, which can be fetched easily, is shown in Table IV, which shows some relations between groups of products, time of sale, temperature, precipitation, and type of customers.

All the linguistic summaries obtained give much insight to the user (analyst) and can be very useful. Moreover, the use of protoforms of linguistic summaries provides a uniformity and can greatly simplify the conceptual and algorithmic design and, hence, the implementation, thus contributing to an inexpensive technology that is so important for a small company.

However, even if the linguistic summaries obtained serve their purpose, they are probably just too simple and too homogenous due to the use of simple and extended protoforms [see (6) and (7)]. The users have welcomed more sophisticated and compound protoforms that could also be, e.g., tailored to their personal preferences or interest. This could be done by developing a more powerful software, but the cost might be prohibitive. However, by using an already-available software for NLG that would accommodate, on one hand, sentence planning or maybe even surface realization and—on the other hand—(simple) phrase-based systems, we would be able to implement, in an easier and less-expensive way, this natural desire for different types of linguistic summaries.

VI. CONCLUSION

We pointed out relevant issues that are related to CWW and argued for an urgent need for a new look at CWW that goes beyond the traditional way of thinking, which has resulted in many very valuable theoretical research results, yet has not led to a proper exposure and recognition in other communities where CWW really belongs and can contribute, i.e., in broadly perceived NLP, notably NLU, and NLG. First, we showed the essence of CWW. In particular, we discussed Zadeh's protoforms, indicated their power, and mentioned a need to develop new tools to handle more modalities. We argued that CWW is not only a powerful conceptual paradigm but also had a high implementational potential. This claim was justified by linguistic data(base) summaries. They are very intuitive and human-consistent natural-language-based knowledge discovery tool and are one of the best examples of the conceptual power of CWW. We first presented their extension and then advocated the use of Zadeh's protoforms as general forms of linguistic data summaries. We showed how fuzzy queries are related to linguistic summaries and that one could introduce, and take advantage of, a hierarchy of protoforms.

In the main part of the paper, we discussed a close relation between linguistic summarization and some concepts and tools in NLG, thus analyzing possible common elements and the use of the tools developed, as well as some inherent differences and difficulties. Notably, we indicated a close relation of linguistic summaries to some type of an extended template-based, and even a simple phrase-based, NLG system. We emphasized that since there exist many software packages for NLG, the tasks mentioned could be implementable in practice and, hence, can be decisive for the wide use of CWW tools. An important part of our discussion, and a clear conclusion from our analysis, was that there is an urgent need to develop new protoforms, going beyond the classic Zadeh's ones, to make full use of the potential of NLG.

For illustration, we showed an implementation for a sales database of a computer retailer, which employs some types of Zadeh's protoforms of linguistic summaries, and we again showed that an urgent need for new types of protoforms also exists in this context; if we could use some software for NLG, this task can be more effectively and efficiently fulfilled.

Clearly, although we used linguistic summaries as the vehicle to analyze and prove our claims, we think that our discussion would be valid for CWW in general. Therefore, we hope that this paper—which presents our personal view and perspective resulting from our long-time involvement in both theoretical work in the area of broadly perceived CWW and a successful real-world implementation—will trigger a discussion and much needed research efforts to help find a way out of a strange situation in which, on one hand, one can clearly see that CWW is related to words (language) and computing and, hence, should be part of huge, respected, and rapidly developing areas of, say, NLP. This area, although it has produced great theoretical results and real-world applications, suffers from a lack of effective and efficient tools for handling imprecision of natural language, which CWW provides. Yet, CWW is practically unknown to that community and not mentioned or cited, even if many authors who work in CWW have already obtained results that could be extremely relevant to that area. On the other hand, even the top people in CWW make no reference to what has been obtained in other natural-language-oriented areas of computer science. This is clearly an example of a wrong and dangerous fragmentation of science. We hope that our paper will play some role to bridge this gap for the benefit of both areas.

REFERENCES

[1] J. Kacprzyk and S. Zadrożny, "FQUERY for access: Fuzzy querying for a Windows-based DBMS," in *Fuzziness in Database Management Systems*, P. Bosc and J. Kacprzyk, Eds. Heidelberg, Germany: Physica-Verlag, 1995, pp. 415–433.

[2] J. Kacprzyk and S. Zadrożny, "CWW in intelligent database querying: Standalone and Internet-based applications," *Inf. Sci.*, vol. 34, pp. 71–109, 2001.

[3] J. Kacprzyk and S. Zadrożny, "Linguistic database summaries and their protoforms: Towards natural language based knowledge discovery tools," *Inf. Sci.*, vol. 173, no. 4, pp. 281–304, 2005.

[4] J. M. Mendel, "The perceptual computer: An architecture for CWW," in *Proc. FUZZ-IEEE*, 2001, pp. 35–38.

[5] J. M. Mendel, "An architecture for making judgement using computing with words," *Int. J. Appl. Math. Comput. Sci.*, vol. 12, no. 3, pp. 325–335, 2002.

[6] J. M. Mendel, "CWW and its relationships with fuzzistics," *Inf. Sci.*, vol. 177, no. 4, pp. 988–1006, 2007.

[7] J. M. Mendel and D. Wu, "Perceptual reasoning: A new CWW engine," in *Proc. GrC*, 2007, pp. 446–451.

[8] I. B. Türksen, "Type 2 representation and reasoning for CWW," *Fuzzy Sets Syst.*, vol. 127, pp. 17–36, 2002.

[9] I. B. Türksen, "Meta-linguistic axioms as a foundation for computing with words," *Inf. Sci.*, vol. 177, no. 2, pp. 332–359, 2007.

[10] V. Novák, "A comprehensive theory of trichotomous evaluative linguistic expressions," *Fuzzy Sets Syst.*, vol. 159, no. 22, pp. 2939–2969, 2008.

[11] E. Trillas, "On the use of words and fuzzy sets," *Inf. Sci.*, vol. 176, no. 11, pp. 1463–1487, 2006.

[12] E. Trillas and S. Guadarrama, "What about fuzzy logic's linguistic soundness?" *Fuzzy Sets Syst.*, vol. 156, no. 3, pp. 334–340, 2005.

[13] F. Herrera and L. Martínez, "A 2-tuple fuzzy linguistic representation model for computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 6, pp. 746–752, 2000.

[14] F. Herrera and L. Martínez, "A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic context in multi-expert decision making," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 31, no. 2, pp. 227–234, Apr. 2001.

[15] F. Herrera and E. Herrera-Viedma, "Linguistic decision analysis: Steps for solving decision problems under linguistic information," *Fuzzy Sets Syst.*, vol. 115, pp. 67–82, 2000.

[16] F. Herrera and E. Herrera-Viedma, "Computing with words: Foundations and applications," *Int. J. Uncertain., Fuzziness Knowl.-Based Syst.*, vol. 9, pp. 1–2, Sep. 2001.

[17] Y. Cao, M. Ying, and G. Chen, "Retraction and generalized extension of computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 6, pp. 1238–1250, Dec. 2007.

[18] J. M. Mendel, "Computing with words: Zadeh, Turing, Popper and Occam," *IEEE Comput. Intell. Mag.*, vol. 2, no. 4, pp. 10–17, Nov. 2007.

[19] F. Liu and J. M. Mendel, "Encoding words into interval type-2 fuzzy sets using an interval approach," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1503–1521, Dec. 2008.

[20] J. Lawry, "A methodology for CWW," *Int. J. Approx. Reason.*, vol. 28, pp. 51–89, 2001.

[21] J. Lawry, "A framework for linguistic modeling," *Artif. Intell.*, vol. 155, no. 1/2, pp. 1–39, 2004.

[22] S. H. Rubin, "Computing with words," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 29, no. 4, pp. 518–524, Aug. 1999.

[23] M. Ying, "A formal model of computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 5, pp. 640–652, Oct. 2002.

[24] H. Wang and D. Qiu, "Computing with words via Turing machines: A formal approach," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 6, pp. 742–753, Dec. 2003.

[25] L. A. Zadeh and J. Kacprzyk, Eds., *CWW in Information/Intelligent Systems, vol. 1., Foundation; vol. 2, Applications*. Heidelberg, Germany: Physica-Verlag, 1999.

[26] D. Wu and J. M. Mendel, "Perceptual reasoning for perceptual computing: A similarity-based approach," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 6, pp. 1397–1411, Sep. 2009.

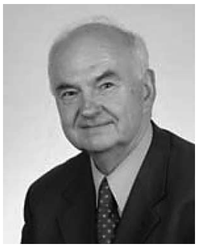
[27] J. Kacprzyk and S. Zadrożny, "Protoforms of linguistic database summaries as a human consistent tool for using natural language in data mining," *Int. J. Softw. Sci. Comput. Intell.*, vol. 1, no. 1, pp. 100–111, 2009.

[28] J. Kacprzyk and S. Zadrożny, "The paradigm of CWW in intelligent database querying," in *CWW in Information/Intelligent Systems*. Heidelberg, Germany: Physica-Verlag, 1999, pp. 382–398.

[29] J. Kacprzyk and S. Zadrożny, "On combining intelligent querying and data mining using fuzzy logic concepts," in *Recent Research Issues on the Management of Fuzziness in Databases*, G. Bordogna and G. Pasi, Eds. Heidelberg, Germany: Physica-Verlag, 2000, pp. 67–81.

[30] J. Kacprzyk and S. Zadrożny, "Data mining via fuzzy querying over the Internet," in *Knowledge Management in Fuzzy Databases*, O. Pons,

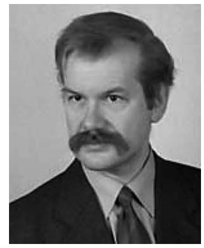
M. A. Vila, and J. Kacprzyk, Eds. Heidelberg, Germany: Physica-Verlag, 2000, pp. 211–233.
[31] J. Kacprzyk and S. Zadrożny, "On a fuzzy querying and data mining interface," *Kybernetika*, vol. 36, pp. 657–670, 2000.
[32] J. Kacprzyk and S. Zadrożny, "On linguistic approaches in flexible querying and mining of association rules," in *Flexible Query Answering Systems. Recent Advances*, H. L. Larsen, J. Kacprzyk, S. Zadrożny, T. Andreasen, and H. Christiansen, Eds. Heidelberg, Germany: Springer-Verlag, 2001, pp. 475–484.
[33] J. Kacprzyk and S. Zadrożny, "Data mining via linguistic summaries of databases: An interactive approach," in *A New Paradigm of Knowledge Engineering by Soft Computing*, L. Ding, Ed. Singapore: World Scientific, 2001, pp. 325–345.
[34] J. Kacprzyk and S. Zadrożny, "Fuzzy linguistic summaries via association rules," in *Data Mining and Computational Intelligence*, A. Kandel, M. Last, and H. Bunke, Eds. Heidelberg, Germany: Physica-Verlag/Springer-Verlag, 2001, pp. 115–139.
[35] J. Kacprzyk and S. Zadrożny, "Linguistically quantified propositions for consensus reaching support," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Budapest, Hungary, Jul. 25–29, 2004, pp. 1135–1140.
[36] J. Kacprzyk and S. Zadrożny, "Supporting consensus reaching in a group via fuzzy linguistic data summaries," in *Fuzzy Logic, Soft Computing and Computational Intelligence (IFSA)*. Beijing, China: Tsinghua Univ. Press/Springer-Verlag, 2005, pp. 1746–1751.
[37] L. A. Zadeh, "Fuzzy logic = CWW," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 103–111, May 1996.
[38] L. A. Zadeh and J. Kacprzyk, Eds., Fuzzy Logic for the Management of Uncertainty. New York: Wiley, 1992.
[39] M. Dertouzos, *The Unfinished Revolution: Human-Centered Computers and What They Can Do for Us*. New York: Harper Collins, 2001.
[40] L. A. Zadeh, "Outline of a new approach to the analysis of complex system and decision processes," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 1, pp. 28–44, Jan. 1973.
[41] L. A. Zadeh, "PRUF—A meaning representation language for natural languages," *Int. J. Man–Mach. Stud.*, vol. 10, pp. 395–460, 1978.
[42] L. A. Zadeh, "Precisiated natural language (PNL)," *AI Mag*, vol. 25, no. 3, pp. 74–91, 2004.
[43] K. von Fintel, "Modality and language," in *Encyclopedia of Philosophy*, 2nd ed. D. M. Borchert, Ed. Detroit, MI: MacMillan, 2006.
[44] R. R. Yager, "A new approach to the summarization of data," *Inf. Sci.*, vol. 28, pp. 69–86, 1982.
[45] R. R. Yager, "On linguistic summaries of data," in *Knowledge Discovery in Databases*, G. Piatetsky-Shapiro and B. Frawley, Eds. Cambridge, MA: MIT Press, 1991, pp. 347–363.
[46] R. R. Yager, "Linguistic summaries as a tool for database discovery," in *Proc. FUZZ-IEEE/IFES*, Yokohama, Japan, 1995, pp. 79–82.
[47] R. R. Yager, "Database discovery using fuzzy sets," *Int. J. Intell. Syst.*, vol. 11, pp. 691–712, 1996.
[48] J. Kacprzyk and R. R. Yager, "Linguistic summaries of data using fuzzy logic," *Int. J. Gen. Syst.*, vol. 30, pp. 133–154, 2001.
[49] J. Kacprzyk, R. R. Yager, and S. Zadrożny, "A fuzzy logic based approach to linguistic summaries of databases," *Int. J. Appl. Math. Comput. Sci.*, vol. 10, pp. 813–834, 2000.
[50] S. Zadrożny and J. Kacprzyk, "Fuzzy querying using the 'query-by-example' option in a Windows-based DBMS," in *Proc. 3rd Eur. Congr. Intell. Tech. Soft Comput.—EUFIT*, Aachen, Germany, vol. 3, 1995, pp. 733–736.
[51] J. Kacprzyk and S. Zadrożny, "Using fuzzy querying over the Internet to browse through information resources," in *Computational Intelligence in Theory and Practice*, B. Reusch and K.-H. Temme, Eds. Heidelberg, Germany: Physica-Verlag, 2001, pp. 235–262.
[52] J. Kacprzyk and S. Zadrożny, "Protoforms of linguistic data summaries: Towards more general natural-language-based data mining tools," in *Soft Computing Systems*, A. Abraham, J. Ruiz del Solar, and M. Koeppen, Eds. Amsterdam, The Netherlands: IOS, 2002, pp. 417–425.
[53] J. Kacprzyk and S. Zadrożny, "Linguistic summarization of data sets using association rules," in *Proc. FUZZ-IEEE*, St. Louis, MO, 2003, pp. 702–707.
[54] R. George and R. Srikanth, "Data summarization using genetic algorithms and fuzzy logic," in *Genetic Algorithms and Soft Computing*, F. Herrera and J. L. Verdegay, Eds. Heidelberg, Germany: Physica-Verlag, 1996, pp. 599–611.
[55] P. Bosc, D. Dubois, O. Pivert, H. Prade, and M. de Calmes, "Fuzzy summarization of data using fuzzy cardinalities," in *Proc. IPMU*, Annecy, France, 2002, pp. 1553–1559.
[56] D. Dubois and H. Prade, "Gradual rules in approximate reasoning," *Inf. Sci.*, vol. 61, pp. 103–122, 1992.
[57] G. Raschia and N. Mouaddib, "SAINTETIQ: A fuzzy set-based approach to database summarization," *Fuzzy Sets Syst.*, vol. 129, pp. 137–162, 2002.
[58] D. Rasmussen and R. R. Yager, "Finding fuzzy and gradual functional dependencies with SummarySQL," *Fuzzy Sets Syst.*, vol. 106, pp. 131–142, 1999.
[59] A. Niewiadomski, "A type-2 fuzzy approach to linguistic summarization of data," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 1, pp. 198–212, Feb. 2008.
[60] Y.-Ch. Hu, R.-Sh. Chen, and G.-H. Tzeng, "Mining fuzzy association rules for classification problems," *Comput. Ind. Eng.*, vol. 43, pp. 735–750, 2002.
[61] J.-H. Lee and H. Lee-Kwang, "An extension of association rules using fuzzy sets," in *Proc. 7th IFSA World Congr.*, Prague, Czech Republic, 1997, pp. 399–402.
[62] J. Kacprzyk and S. Zadrożny, "Fuzzy querying for microsoft access," in *Proc. FUZZ-IEEE*, Orlando, FL, vol. 1, 1994, pp. 167–171.
[63] J. Kacprzyk and S. Zadrożny, "Fuzzy queries in Microsoft Access v. 2," in *Proc. FUZZ-IEEE/IFES, Workshop Fuzzy Database Syst. Inf. Retrieval*, Yokohama, Japan, 1995, pp. 61–66.
[64] S. Zadrożny, G. De Tre, R. De Caluwe, and J. Kacprzyk, "An overview of fuzzy approaches to flexible database querying," in *Handbook of Research on Fuzzy Information Processing in Databases*, J. Galindo, Ed. Hershey, PA: Idea Group, 2008, pp. 34–53.
[65] L. A. Zadeh, "A prototype-centered approach to adding deduction capabilities to search engines—The concept of a protoform," presented at the BISC Semin., Berkeley, CA, 2002.
[66] L. A. Zadeh, "A computational approach to fuzzy quantifiers in natural languages," *Comput. Math. Appl.*, vol. 9, pp. 149–184, 1983.
[67] L. A. Zadeh, "Syllogistic reasoning in fuzzy logic and its application to usuality and reasoning with dispositions," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 6, pp. 754–763, Nov./Dec. 1985.
[68] R. R. Yager, "On ordered weighted averaging operators in multicriteria decision making," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-18, no. 1, pp. 183–190, Jan./Feb. 1988.
[69] R. R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators: Theory and Applications*. Boston, MA: Kluwer, 1997.
[70] S. Zadrożny and J. Kacprzyk, "Summarizing the contents of Web server logs: A fuzzy linguistic approach," in *Proc. FUZZ-IEE*, London, U.K., 2007, pp. 100–105.
[71] J. Kacprzyk, A. Wilbik, and S. Zadrożny, "On some types of linguistic summaries of time series," in *Proc. 3rd Int. IEEE Conf. Intell. Syst.*, London, U.K., Sep. 4–6, 2006, pp. 373–378.
[72] J. Kacprzyk, A. Wilbik, and S. Zadrożny, "Linguistic summarization of time series under different granulation of describing features," in *Lecture Notes in Artificial Intelligence*. Heidelberg, Germany: Springer-Verlag, vol. 4585, 2007, pp. 230–240.
[73] J. Kacprzyk, S. Zadrożny, and A. Wilbik, "Linguistic summarization of some static and dynamic features of consensus reaching," in *Computational Intelligence, Theory and Applications*, B. Reusch, Ed. Berlin, Germany: Springer-Verlag, 2006, pp. 19–28.
[74] J. Kacprzyk and A. Ziółkowski, "Database queries with fuzzy linguistic quantifiers," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-16, no. 3, pp. 474–479, May 1986.
[75] J. Kacprzyk, S. Zadrożny, and A. Ziółkowski, "FQUERY III+: A 'human consistent database querying system based on fuzzy logic with linguistic quantifiers," *Inf. Syst.*, vol. 6, pp. 443–453, 1989.
[76] J. Kacprzyk, G. Pasi, P. Vojtaš, and S. Zadrożny, "Fuzzy querying: Issues and perspective," *Kybernetika*, vol. 36, pp. 605–616, 2000.
[77] L. A. Zadeh, "From search engines to question answering systems—The problems of world knowledge relevance deduction and precisiation," in *Fuzzy Logic and the Semantic Web*, E. Sanchez, Ed. Amsterdam, The Netherlands: Elsevier, 2006, pp. 163–210.
[78] S. Sripada, E. Reiter, and I. Davy, "SumTime-Mousam: Configurable marine weather forecast generator," *Expert Update*, vol. 6, no. 3, pp. 4–10, 2003.
[79] F. Portet, E. Reiter, A. Gatt, J. Hunter, S. Sripada, Y. Freer, and C. Sykes, "Automatic generation of textual summaries from neonatal intensive care data," *Artif. Intell.*, vol. 173, pp. 789–816, 2009.
[80] E. Reiter and R. Dale, *Building Natural Language Generation Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[81] J. Kacprzyk and S. Zadrożny, "Soft computing and Web intelligence for supporting consensus reaching," *Soft Comput.*, to be published. DOI: 10.1007/s00500-009-0475-4.



Janusz Kacprzyk (M'91–SM'01–F'06) received the M.Sc. degree in automatic control from the Warsaw University of Technology, Warsaw, Poland, the Ph.D. degree in systems analysis, and the D.Sc. ("habilitation") degree in computer science from the Systems Research Institute, Polish Academy of Sciences, Warsaw.

He was a Visiting Professor in the United States, Italy, and China. He is currently a Professor of computer science with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland. He is also a Professor with Warsaw School of Information Technology and with the Industrial Research Institute for Automation and Measurements, Warsaw. His research interests include soft computing, fuzzy logic, and computing with words, in decisions and optimization, control, database querying, and information retrieval. He is the author of five books, the editor or coeditor of 30 volumes, and the author or coauthor of 300 papers. He is also the Editor-in-Chief of four book series at Springer and two journals.

Prof. Kacprzyk is a member of the Polish Academy of Sciences and the Spanish Royal Academy of Economic and Financial Sciences. He is a Fellow of the IFSA. He has received many awards, including the 2005 IEEE Computational Intelligence Society Pioneer Award in Fuzzy Systems, the Sixth Kaufmann Prize, and the Gold Medal for pioneering works on soft computing in economics. He is currently the President of the Polish Society for Operational and Systems Research and the Immediate Past President of the International Fuzzy Systems Association (IFSA).



Sławomir Zadrożny received the M.Sc. degree in computer science from the University of Warsaw, Warsaw, Poland, and the Ph.D. and D.Sc. ("habilitation") degrees in computer science from the Systems Research Institute, Polish Academy of Sciences, Warsaw.

He is currently an Associate Professor with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland. He is also an Associate Professor with Warsaw School of Information Technology and with the Technical University of Radom, Radom, Poland. His research interests include applications of fuzzy logic in database-management systems, data mining, information retrieval, and decision support. He is the author or coauthor of approximately 150 journal and conference papers, two books, several edited volumes, and special issues of scientific journals. He has been involved in the design and implementation of several prototype software packages.