UNECE

CONFERENCE OF EUROPEAN STATISTICIANS

Workshop on Statistical Dissemination and Communication WP.4-2

12-14 June 2019, Gdansk, Poland 07 June 2019

Disseminating Statistical Data by Short Quantified Sentences of Natural Language

Miroslav Hudec (University of Economics in Bratislava, Slovakia)

miroslav.hudec@euba.sk & hudec@fon.bg.ac.rs

Abstract and Paper

One of the key missions of national and international statistical institutes is the dissemination of statistical data to diverse data users. Data from these institutes are generally considered an important source of credible evidence. Summarization and dissemination via traditional methods is a convenient approach for providing this evidence. However, this is usually comprehensible only for users with a considerable level of statistical literacy. Kahneman (2011) observed that people are generally good intuitive grammarians, but they have not a similar intuitive feel for the key principles of statistics. Thus, National Statistical Institutes (NSIs) should apply different strategies in order to meet the expectations of diverse user categories (businesses, public administrations, researchers, journalists, and the general public) who are increasingly interested in information that describe various aspects of our society. Thus, statistical agencies should offer flexibility in dissemination to avoid jeopardizing their mission (Bavdaž 2011). This may require rules for using natural human languages to describe the key measures (Schield 2011) and to make statistics easily understandable and usable by the general public (Bier and Nymand-Andersen 2011).

A promising alternative is augmenting the data dissemination and statistical figures linguistically (Hudec at al 2018). Especially, the less statistically literate users (e.g., domain experts and the general public) as well as disabled people can benefit from such summarization. This work studies the potential of explaining data summaries by short quantified sentences of natural language. Summaries like "most visits from remote countries are of a short duration" or "few observations are around the mean value" can be immediately understood by diverse data users. Linguistic summaries are not intended to replace existing dissemination approaches, but can augment them by providing alternatives for the benefit of data users of official statistics. To reflect the elasticity of adverbs, adjectives and verbal quantifiers, fuzzy sets offer a "sliding scale" definition of linguistic terms by membership functions (Hudec 2016, Lesot et al 2016, Liu 2011, Zadeh 1996). To avoid summaries based on the outliers, or low data coverage, a quality criterion is suggested (Hudec 2017). The potential of linguistic data summaries is demonstrated on illustrative interfaces on real-world data. In order to reduce both: the complexity and requirements on users to intervene, we have suggested an approach that minimally burden users. Furthermore, this approach could support automated or computational journalism (Coddington 2015, Graefe 2016) which might be welcome for journalists

searching for statistical information to support their articles (while the data journalists would prefer the full access to the data).

Finally, our research has documented perspectives, obstacles and problems leading to the future research directions to develop broadly accepted designs for the full-featured and easy-to-use interfaces. These tasks should be solved in the cooperation between NSI data dissemination units, and scientists and practitioners working in the various fields.

Based mainly on the paper "Augmenting Statistical Data Dissemination by Short Quantified Sentences of Natural Language" Journal of Official Statistics, Vol. 34, No. 4, 2018, pp. 981–1010.

Main references

Bavdaž M. (editor). (2011) Final Report Integrating Findings on Business Perspectives Related to Nsis' Statistics. Brussels: European Commission. (Deliverable 3.2 from FP7 project BLUE-Enterprise and Trade Statistics.

Bier V., Nymand-Andersen P. (2011) Communicating Statistics to Frequent Users – One Size Fits All? In Proceedings of the Committee for the Coordination of Statistical Activities (CCSA Special Session), September 8, 2011. Luxembourg.

Coddington M. (2015) Clarifying Journalism's Quantitative Turn. *Digital Journalism*, 3:331–348.

Graefe A. (2016) *Guide to Automated Journalism*. Tow Center for Digital Journalism, New York.

Hudec M. (2016) *Fuzziness in Information Systems - How to Deal with Crisp and Fuzzy Data in Selection, Classification, and Summarization*. Springer, Cham.

Hudec M. (2017) Merging Validity and Coverage for Measuring Quality of Data Summaries. In *Information Technology and Computational Physics*, edited by P. Kulczycki, L.T. Kóczy, R. Mesiar, and J. Kacprzyk, 71–85. Springer, Cham.

Hudec M., Bednárová E., Holzinger A. (2018) Augmenting Statistical Data Dissemination by Short Quantified Sentences of Natural Language. *Journal of Official Statistics,* 34(4):981–1010.

Kahneman D. (2011) Thinking Fast and Slow. Farrar, Staraus and Giroux, New York.

Lesot M-J., Moyse G., Bouchon-Meunier B. (2016) Interpretability of Fuzzy Linguistic Summaries. *Fuzzy Sets and Systems*, 292:307–317.

Liu B. (2011) Uncertain Logic for Modeling Human Language. *Journal of Uncertain Systems*, 5:3–20.

Schield M. (2011) Statistical Literacy: A New Mission for Data Producers. Statistical Journal of the IAOS 27:173–183.

Zadeh L.A. (1996) Fuzzy logic = computing with words. *IEEE transactions on fuzzy systems*, 4(2):103-111.

UNECE - Conference of European Statisticians
Workshop on Statistical Dissemination and Communication
12 -14 June 2019, Gdańsk, Poland

Disseminating Statistical Data by Short Quantified Sentences of Natural Language

Miroslav Hudec
(Faculty of Economic Informatics, University of Economics in Bratislava, Slovakia, miroslav.hudec@euba.sk)

1 Introduction

One of the key missions of national and international statistical institutes is the dissemination of statistical data to diverse users' categories (e.g., businesses, public administration, researchers, general audience). Data from these institutes are generally considered as an important source of credible evidence. Summarization and dissemination via traditional methods is a convenient approach for providing this evidence. However, this is usually comprehensible only for users with a considerable level of statistical literacy. Kahneman (2011) observed that people are generally good intuitive grammarians, but they have not a similar intuitive feel for the key principles of statistics. Less statistically literate users (e.g., domain experts and the general public) should also benefit from the disseminated data. Thus, statistical agencies should offer flexibility in dissemination to avoid jeopardizing their mission (Bavdaž 2011). This may require rules for using natural human languages to describe key measures (Schield 2011) and to make statistics easily understandable and usable for the general public (Bier and Nymand-Andersen 2011).

The mathematical formalization of a sliding scale definition of linguistic terms (such as *high productivity*, *low productivity* and *around the mean value*) as well as the elasticity of verbal quantifiers related to the relative proportions or frequencies of units possessing these attributes (*few enterprises*, *about half of the enterprises*, *most enterprises* and the like) is realized by fuzzy sets expressed through the membership functions and logical operators (Hudec 2016, Lesot et al. 2016, Liu 2011, Zadeh 1996, Zadeh 1983). Summaries including, e.g., "*most visits from remote countries are of a short duration*" can be immediately understood by diverse users regardless of their statistical literacy. To avoid summaries based on outliers, low data coverage and improperly formalized fuzzy sets, various quality measures have been suggested, e.g., (Hudec 2017; Kacprzyk and Strykowski, 1999, Wu et al. 2010).

These observations lead to the augmenting statistical data dissemination linguistically. In this paper the focus is on the short quantified sentences of natural language proposed by Hudec et al. (2018). In this paper, a theoretical view on dissemination by linguistic summaries supported with the examples on the real statistical data is discussed.

2 Motivation

Inspiration for this work evolved from the following observations: (i) graphical interpretation is a valuable way of summarization; however, it is not always effective (Lesot et al. 2016); (ii)

users (e.g., small businesses) are often interested in summarized information rather than data (Bavdaž 2011); (iii) summaries should not be as terse as means and should hold for any data type and distribution (Yager et al. 1990); (iv) a natural way for humans to communicate, compute and conclude is natural language (Zadeh 2001); and (v) existing approaches, to our best knowledge, in data dissemination are typically based on precise conditions.

The alternative is summaries by short quantified sentences of natural language, or linguistic summaries. For instance, we can say: *the average age is 40.3 with a standard deviation of 17.53*, or: "*few respondents are near the average age*". The second case clearly illustrates that the mean value is not a sufficiently representative characteristic in this example (see Figure 1). People might overlook the dispersion measures and focus only on the measure expressing average (and moreover link all average functions with the arithmetic mean).

Further, we can interpret the summary between attributes, e.g., "*most visits from remote countries are of a short duration*". Such summary, although neither based on traditional mathematical methods nor on visualisation on graph or map, contains very valuable information for domain experts in truism, accommodation providers, marketers, journalists and local authorities.

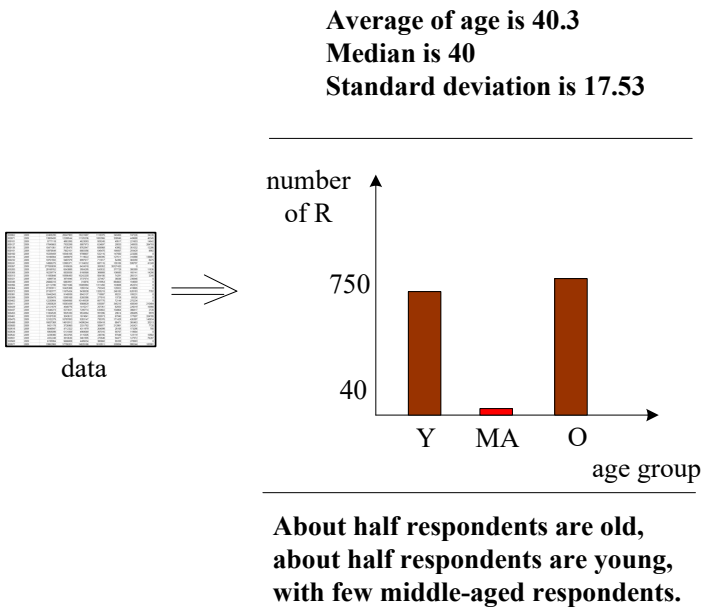

Figure 1: Interpreting data by statistical figures, on graph, and by short quantified sentences.

In Figure 1 we see three ways for interpreting the same information about one attribute. But, for expressing relational knowledge among attributes, e.g., aforementioned correlation between countries and duration of visits, the suitable way is by quantified sentences. Moreover, a short quantified sentence can be interpreted by text-to-speech synthesis systems, a convenient way, whenever the users' visual attention is focused on something else, or for the disabled users.

3 Theoretical preliminaries

The main categories of linguistic summaries are (Lesot et al. 2016): classic prototype forms, prototype forms of time series and temporal prototype forms. The classic prototype forms

summarize the proportion on the whole data set, or on a restricted part. The former case (known also as a basic structure) is illustrated by the sentence "*most houses have high gas consumption"*, whereas an illustrative example of the latter (a summary with restriction) is: "*most of the old houses have high gas consumption"*. The prototype forms of time series linguistically express behaviour of attributes over time. A illustrative sentence is "*about half small businesses have small response rate most of the time"*. The temporal prototype forms do not use verbal quantifiers, but mode of behaviour for creating periodic summaries, e.g., "r*egularly in spring, the participation in surveys takes high value*".

For the illustrative example in Figure 1 (the basic structure of classic prototype form), we used three overlapping fuzzy sets for attribute *age* and *proportion*. Three terms: *young*, *middle-aged* and *old* are formalized as follows

$$\mu_{young}(x) = \begin{cases} 1 & x \leq 30 \\ (35 - x)/5 & 30 < x < 35 \\ 0 & x \geq 35 \end{cases}$$

$$\mu_{middle\_aged}(x) = \begin{cases} (x - 30)/5 & 30 < x < 35 \\ 1 & 35 \leq x \leq 50 \\ (55 - x)/5 & 50 < x < 55 \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_{old}(x) = \begin{cases} 0 & x \leq 50 \\ (x - 50)/5 & 50 < x < 55 \\ 1 & x \geq 55 \end{cases}$$

In this way, similar values might influence two neighbouring sets, but the sum of intensities of belonging to these sets should be equal to 1, i.e. a respondent aged 25 fully influences the size or cardinality of set *young*, whereas a respondent aged 32 influences the set *young* with degree of 0.6 and the set *middle-aged* with degree of 0.4. The family of verbal relative quantifiers is shown in Figure 2.

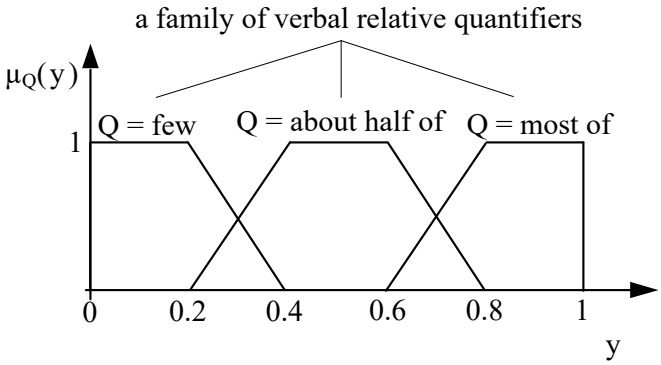

Figure 2: A possible family of relative quantifiers, where *y* is a proportion of respondents which fully or partially belong to concepts expressed by fuzzy sets, in our case, *young*, *middle-aged* and *old*.

We should emphasize, that the parameters of fuzzy sets are context dependent. They might vary, but the shape remains the same. Sets like *young*, *short distance*, *short duration*, *low pollution* are always monotone decreasing, whereas sets like *old*, *high distance*, *long duration*, *high pollution* ale always monotone increasing functions.

The validity for the basic structure is computed in the following way (Yager 1982)

$$v(Qx(Px)) = \mu_Q(\frac{1}{n}\sum_{i=1}^{n}\mu_S(x_i)) \quad (1)$$

where n is the number of tuples (records) in a data set (cardinality), $\frac{1}{n}\sum_{i=1}^{n}\mu_S(x_i)$ is the proportion of tuples in a data set that satisfy concept S and μ_Q is the membership function of chosen relative quantifier (Figure 2). Obviously, the validity assumes values from the unit interval.

The validity for summaries with restriction is computed in the following way (Rasmussen and Yager 1997):

$$v(Qx(Px)) = \mu_Q(\frac{\sum_{i=1}^{n}(\mu_S(x_i) \wedge \mu_R(x_i))}{\sum_{i=1}^{n}\mu_R(x_i)}) \quad (2)$$

where $\frac{\sum_{i=1}^{n} t(\mu_S(x_i),\mu_R(x_i))}{\sum_{i=1}^{n}\mu_R(x_i)}$ is the proportion of tuples in a data set that satisfy concept S and belong to restricted part R, and μ_Q is the membership function of chosen relative quantifier.

Regarding the example shown in Figure 1, the validates for all nine sentences are

- Most of commuters are young 0
- Most of commuters are middle aged 0
- Most of commuters are old 0
- **About half of commuters are young 0.8570**
- About half commuters are middle aged 0.1425
- **About half commuters are old 1**
- Few customers are young 0.1430
- **Few customers are middle aged 0.8575**
- Few customers are old 0

The sentences marked as bold have significant validities and therefore are displayed in Figure 1, whereas the other six have low validities and thus have failed a filter.

Another important question is the quality of summary. To avoid summaries based on the low validity, low data coverage, outliers, improperly constructed fuzzy sets, etc. the literature offers a higher number of quality measures tailored to the structures of summaries. A brief overview is in, e.g., (Hudec et al. 2018). A simplified quality criterion for structures with the restriction (a conjunction of the sufficient data coverage validity) is proposed by Hudec (2017). This quality measure is integrated into the quality of summary examined in the third example bellow.

4 Illustrative examples

The following three examples illustrate this idea on the municipal statistics database. The procedures are based on SQL-like queries empowered with the fuzzy logic calculus.

4.1 The distribution around the mean value

Let us recall an example from Hudec et al. (2018). A historian wishes to examine the mean value of the *year of the first written notice* (an attribute in the Slovak municipal statistics database – approx. 800 indicators for 2,927 municipalities). The historian has divided municipalities into two sets: "*population less than 12,000*" and "*population greater than or equal to 12,000*". The interface for interpreting solutions is shown in Figure 3. In this interface, the user can choose the relevant attribute from the database and relative dispersion around the mean value, which is set to 10% by default.

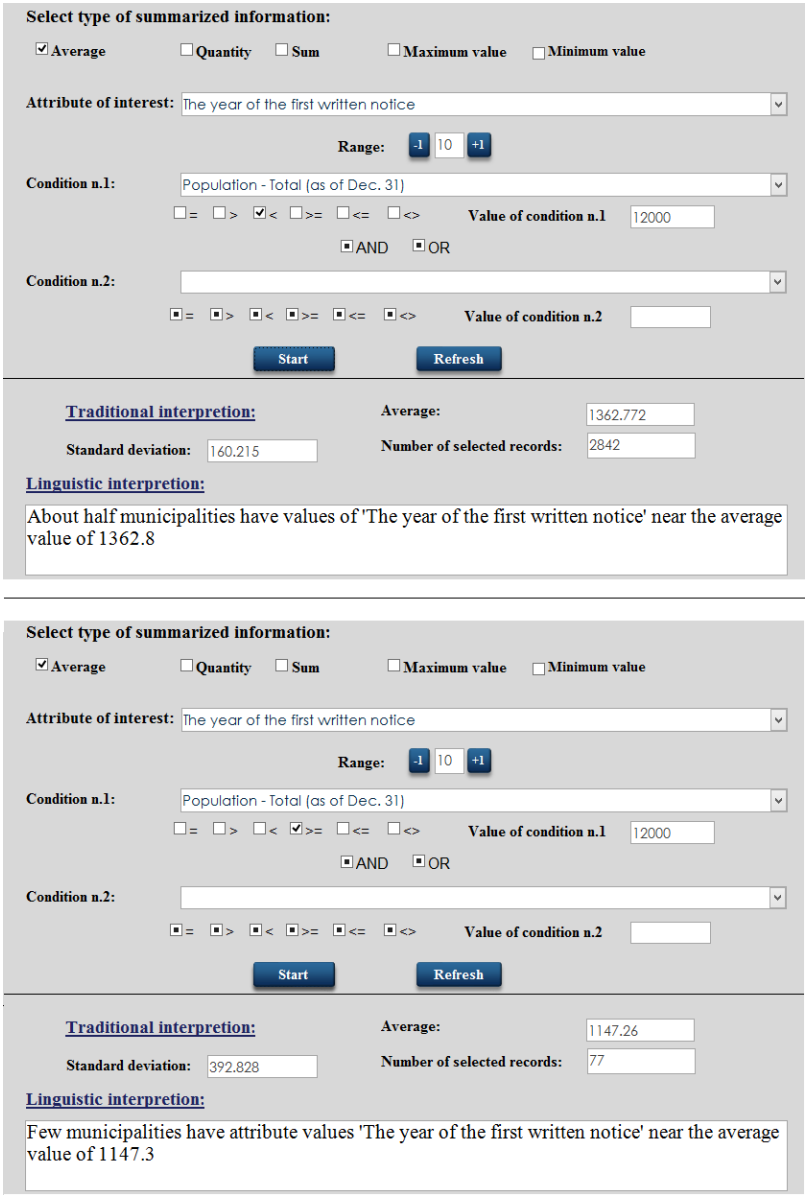

Figure 3: The interfaces for linguistically interpreting data distribution around the mean value (Hudec et al. 2018).

Via this interface, the user discovers that the mean value of *the year of first written notice* for the municipalities with low population is the year 1363 as well as that about half of them have their year of first written notice in the vicinity of the mean value (Figure 3 – upper interface). Hence, the arithmetic mean is a suitable generalization. For the municipalities with high population, the situation is the opposite. The mean value is the year 1147, but only few municipalities fully or partially belong to the neighbourhood of this value (Figure 3 – lower interface).

This interpretation is suitable for diverse users' categories, because the well-known statistical measures are disseminated together with their verbal interpretations.

4.2 Quantified sentence as a nested query condition

The nested query conditions expressed by linguistic summaries can be applied on hierarchical data structures, e.g., territorial levels expressed by NUTS (Nomenclature of Territorial Units for Statistics) and LAU (Local Administrative Units). For instance, an enterprise might be interested for regions, where most of municipalities meet particular condition expressed linguistically. The procedure is roughly the same as for the other examples. The main difference is that the procedure is realized for each unit on the higher level. For instance, an enterprise is interested in extending its business activities related to the agriculture into the areas of low altitude. Hence, the question is: "*find regions where most of the municipalities have low altitude*". The answer is regions ranked downward according to the intensity of satisfying the summarized sentence shown in Table 1. The remaining two regions are not selected, because they do not meet the query condition. A short glance at a map provides similar conclusion.

Table 1: Retrieved and ranked regions.

Region	**Rank**
Bratislava	1
Trnava	1
Nitra	1
Trenčín	0.7719
Košice	0.6314
Bánska Bystrica	0.2116

The condition can be straightforwardly extended with several logically connected conditions like "*find regions where most of the municipalities have high altitude and low pollution and low population density*".

Generally, validates of summaries might be depicted on a thematic map. For instance, territorial units which fully satisfy the requirement can be marked with some colour, territorial units which do not satisfy the query can be marked with another colour (quite different than the one for expressing validity equal to 1) and territorial units which partially meet the query condition are marked again with some other colour having a gradient from faint hue to deep hue according to their validities.

4.3 Summaries between attributes

The classic prototype form of summary with restriction covers this requirement. For instance, a demographer is interested to learn, whether *"the majority of municipalities with a high ratio of arable land have a low population density"*. Therefore, the validity of summary among attributes should be calculated. The interface is shown in Figure 4. The user can directly assign values to the respective parameters of linguistic terms, or ask the system for suggestions. In the latter case, parameters are mined from the database and presented to the user who has the choice to either accept or modify them. The validity of this summary is 0.814 (by Eq. 2) and its quality is 0.8082 (by Eq. from (Hudec 2017)). These high values lead us to the conclusion that the summarized sentence is of a high quality. The intensities of validity and quality are visualized on sliders to indicate how far from the ideal values they are. This interface might be tedious for less skilled users, but advanced users might benefit from the offered possibilities.

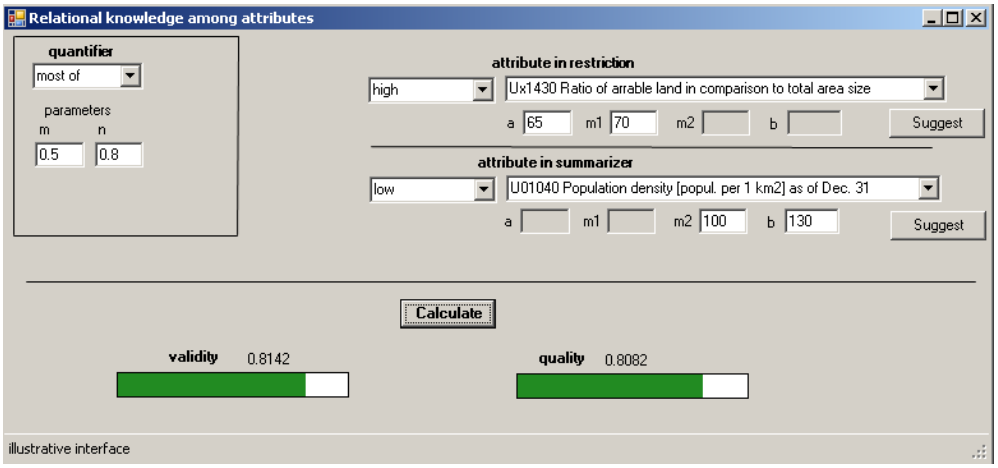

Figure 4: The illustrative interface for analysing verbal summaries with restriction (Hudec et al. 2018).

The interfaces in Figures 3 and 4 are not the full-featured ones. The technique has been developed into illustrative interfaces to demonstrate the applicability.

5 The main features and possible applicability

The main features of the suggested approach are the following (Hudec et al. 2018):

- The less complex interpretation of the data is especially welcome for less statistically literate users and disabled people, for whom the summarized sentences may be interpreted by voice.
- This way can easily be applied to any human language. Concepts such as *high* and verbal quantifiers such as *most of* are always expressed by increasing functions, regardless of their translation to the other languages.
- It is less sensitive to the imprecise nature of some data. When the measured value is not far from the real one, then this approach eliminates sharp jumps between belonging and not belonging to a concept.
- The suggested approach reveals summaries from the data, not the data itself. Generally speaking, the data disclosure would not be a problem; however, care should be taken

when summarizing from small data sets. Thus, the decision on which data sources might be available for users to realize summaries should meet regulations and other relevant rules.

- LSs are able to offer an alternative answer when the initial sentence (summary) is of insufficient validity or quality. For instance, if the proportion for the sentence "*most short visits are from countries with high GDP"* is 0.055, the answer is not only that the validity is zero, but we can provide an alternative summary: "*few short visits are from countries with high GDP"*.
- Statistical offices typically refrain from disseminating dispersion measures, although this information is valuable. The proposed approach includes the linguistically interpreted deviation and distribution, which is suitable for all users, especially for the less statistically literate ones.

The question is a practical realization for diverse users' groups. Regarding the data users which are not obliged to cooperate in surveys and occasionally explore disseminated information, statistical institutes might provide simpler interfaces. Concerning data users which have to frequently cooperate in data collection, the situation is different. It was estimated by Adolfsson et al. (2010) that 30% of total data collection costs is allocated to data imputation. Ross (2009) observed the still valid paradox that users of official statistics are becoming more demanding for data, but are less willing to provide their own data. Although data collection and dissemination are at two opposite ends of the statistical data production process, they are influencing each other. As a motivation, statistical institutes might offer sophisticated methods for linguistically interpreted summaries (means, deviations, time series, mode of behaviour, etc.) to business' respondents that cooperate timely in surveys. By this way we might mitigate the aforementioned paradox. The practical feasibility of achieving this (while maintaining the principle of impartiality) is a topic for future research.

Linguistic summaries might be supportive for the emerging field of automated or computational journalism, i.e. technologically oriented journalism focused on the application of computational intelligence to the practices of information gathering and information presentation (Coddington 2015, Graefe 2016). This approach might be welcome for journalists searching for aggregated statistical information to support their articles.

A possible obstacle might be the structure of short quantified sentences. The order of terms and the structure itself might not fully meet the usual terminology in official statistics and grammar rules. Thus, there is a room for experts from different fields to identify sound and practical solutions. The interactive machine learning could be of help here. Verbal explanations are extremely important for the emerging field of "explainable artificial intelligence" (Goebel et al. 2018). For instance, in example on Figure 4 the user selects which attributes should be considered. On the other hand, a mechanical construction of sentences may lead to the correlation among attributes which does not reflect causality, or for users irrelevant summaries. Thus, the synergy of machine learning and human-in-the-loop is a perspective direction.

6 Concluding remarks

Linguistic summaries play a pivotal role in summarizing information from the data when uncertainty related to the semantic meaning of the phenomena cannot be neglected, or users cannot clearly express their requirements. For the users it is more convenient to explain

questions by linguistic terms rather than numbers. This work has speculated possibilities for applying short quantified sentences in statistical data dissemination, because linguistically summarized information is understandable for a large scale of statistical data users regardless of their statistical literacy.

The important further activity is in developing full-featured and easy-to-use interfaces and in real-world testing with users to develop broadly accepted solutions, which might be further adjusted to the particular cases. These tasks should be solved in cooperation between national statistical institutes' data dissemination units, and scientists and practitioners working in the various fields. In order to bring this approach from the theoretical perspective to the full-working approach, this cooperation is advisable.

Finally, we emphasize that the suggested approach based on linguistic summaries should not be considered as a competitor to existing ones, but rather as a complementary dissemination practice to well-established ones.

References

Adolfsson, C., Arvidson, G., Gidlund, P., Norberg, A., Nordberg, L. (2010) Development and Implementation of Selective Data Editing at Statistics Sweden. In Proceedings of the European Conference on Quality in Official Statistics, May 4, 2010. Helsinki.

Bavdaž M. (editor). (2011) Final Report Integrating Findings on Business Perspectives Related to Nsis' Statistics. Brussels: European Commission. (Deliverable 3.2 from FP7 project BLUE-Enterprise and Trade Statistics.

Bier V., Nymand-Andersen P. (2011) Communicating Statistics to Frequent Users – One Size Fits All? In Proceedings of the Committee for the Coordination of Statistical Activities (CCSA Special Session), September 8, 2011. Luxembourg.

Coddington M. (2015) Clarifying Journalism's Quantitative Turn. Digital Journalism, 3:331–348.

Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., Kieseberg, P., Holzinger, A. (2018) Explainable AI: The New 42? In Machine Learning and Knowledge Extraction, Springer Lecture Notes in Computer Science LNCS 11015, edited by A. Holzinger, P. Kieseberg, A. Tjoa and E. Weippl, 295-303. Cham: Springer.

Graefe A. (2016) Guide to Automated Journalism. Tow Center for Digital Journalism, New York.

Hudec M. (2016) Fuzziness in Information Systems - How to Deal with Crisp and Fuzzy Data in Selection, Classification, and Summarization. Springer, Cham.

Hudec M. (2017) Merging Validity and Coverage for Measuring Quality of Data Summaries. In Information Technology and Computational Physics, edited by P. Kulczycki, L.T. Kóczy, R. Mesiar, and J. Kacprzyk, 71–85. Springer, Cham.

Hudec M., Bednárová E., Holzinger A. (2018) Augmenting Statistical Data Dissemination by Short Quantified Sentences of Natural Language. Journal of Official Statistics, 34(4):981–1010.

Kacprzyk, J., Strykowski, P. (1999) Linguistic Data Summaries for Intelligent Decision Support. In Proceedings of the fourth European Workshop on Fuzzy Decision Analysis and

Recognition Technology for Management, Planning and Optimization (EFDAN 1999), June 14 – 15, 1999. 3-12. Dortmund.

Kahneman D. (2011) Thinking Fast and Slow. Farrar, Staraus and Giroux, New York.

Lesot M-J., Moyse G., Bouchon-Meunier B. (2016) Interpretability of Fuzzy Linguistic Summaries. Fuzzy Sets and Systems, 292:307–317.

Liu B. (2011) Uncertain Logic for Modeling Human Language. Journal of Uncertain Systems, 5:3–20.

Rasmussen, D., Yager, R.R. (1997) Summary SQL - A Fuzzy Tool for Data Mining Intelligent Data Analysis, 1:49-58.

Ross, M.P. (2009) Official Statistics in Malta – Implications of Membership of the European Statistical System for a Small Country/NSI. In Proceedings of the 95th DGINS Conference, October 1, 2009. Malta.

Schield M. (2011) Statistical Literacy: A New Mission for Data Producers. Statistical Journal of the IAOS 27:173–183.

Wu, D., Mendel, J.M., Joo, J. (2010) Linguistic Summarization Using If-Then Rules. In Proceedings of the 2010 IEEE International Conference on Fuzzy Systems, July 18-23, 2010. 1–8. Barcelona.

Yager, R.R. (1982) A new approach to the summarization of data. Information Sciences, 28: 69-86.

Yager, R.R., Ford, M., Canas, A.J. (1990) An Approach to the Linguistic Summarization of Data. In Proceedings of the 3rd International Conference of Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU 1990), July 2-6, 1990. 456–468. Paris.

Zadeh, L.A. (2001) From Computing With Numbers to Computing With Words—From Manipulation of Measurements to Manipulation of Perceptions. In Computing with Words edited by P. Wang, 35–68. Wiley, New York.

Zadeh L.A. (1996) Fuzzy logic = computing with words. IEEE transactions on fuzzy systems, 4(2):103-111.

Zadeh, L.A. (1983) A Computational Approach to Fuzzy Quantifiers in Natural Languages. Computers & Mathematics with Applications, 9:149–184.

Zimmermann, H.J. (2001) Fuzzy Set Theory—and its Applications. Kluwer Academic Publishers, Dordrecht.