

Accessible Numerical Information: Cookery Recipes as a Special Case

Susana Bautista^{#1}, Raquel Hervás^{#2}, Pablo Gervás^{#3}

*#Departamento Ingeniería del Software e Inteligencia Artificial,
Facultad de Informática, Universidad Complutense de Madrid
Profesor José García Santesmases, 9 28040 Madrid, Spain*

¹subautis@fdi.ucm.es

²raquelhb@fdi.ucm.es

³pgervas@sip.ucm.es

Abstract— This paper presents a proposal to solve the need of making numerical information accessible. We use the domain of cooking recipes as a special case of study. For the numerical information in recipes to be accessible it is necessary to adapt the different mathematical forms used for the quantities of different ingredients. It is a first approach to automate the adaptation of the recipes for universal accessibility.

Keywords— numerical information, adaptation, mathematical form, accessible, recipe

I. INTRODUCTION

The problem of making information accessible to all has become an important area of research in recent years. Efforts along these lines include the development of accessible texts following different guidelines as well as efforts to automate this process.

“The Standard Rules on the Equalization of Opportunities for Persons with Disabilities” by United Nations [1] state that all public information services and documents should be accessible in such a way they could be easily understood. If we focus on numerical information, a large percentage of numerical information appears in different kinds of documents like newspapers, household bills and cooking recipes.

However, a lot of people find this type of information difficult to understand, mainly people with poor education or special needs. Sometimes numerical information is presented using a mathematical form difficult to understand for these people, for example using fractions or percentages. At other times numerical information is presented with a modifier to indicate explicitly that some loss of precision has taken place, for example, “around 1 ½ kg”.

In the text simplification process to make information accessible, different tasks are carried out: replacing difficult words, splitting sentences, etc. The simplification of numerical expressions is one of these tasks.

This paper presents a special case of study, that of cookery recipes, where numerical information for ingredients is shown in different mathematical form or using different units of measure, for example, “60 gr sugar”, “1/4 flour” or “0.5 l water”. Our final aim is automate the process of adaptation of difficult numerical expressions in cookery recipes based on

previous conclusions of empirical studies in simplification of numerical expressions [2], [3], [4].

The remainder of the article is organised as follows: in Section II we present an overview of the most relevant work in the field of automatic text simplification focusing on numerical information simplification; in Section III we outline our approach to the task at hand and describe in some detail the analysis of the cookery recipes considered, while Section IV presents our discussion of the proposal. We conclude the article with some conclusions and plans for future work in Section V.

II. PREVIOUS WORK

Text simplification, a relatively new task in Natural Language Processing, has been directed mainly at syntactic constructions and lexical choices that some readers find difficult, such as long sentences, passives, coordinate and subordinate clauses, abstract words, low frequency words, and abbreviations.

The rule-based paradigm has been used in the implementation of some systems for text simplification, each one focusing on a variety of readers (with poor literacy, aphasia, etc) [5], [6], [7], [8].

The transformation of texts into easy-to-read versions can also be phrased as a translation problem between two different subsets of language: the original and the easy-to-read version. Corpus-based systems can learn from corpora the simplification operations and also the required degree of simplification for a given task [9], [10], [11].

A variety of simplification techniques have been used, replacing uncommon words with more common ones [12], activating passive sentences and resolving references [13], reducing multiple-clause sentences to single-clause sentences [14], [13], [15] and making appropriate choices at the discourse level [16]. Reference [17] studied the trade off between brevity and clarity in the context of generating referring expressions. Other researchers have focused on the generation of readable texts for readers with low basic skills [18], and for teaching foreign languages [10].

Previous work on numerical expressions has studied the treatment of numerical information in different areas like health care [19], forecast [20], representation of probabilistic

information [21] or vague information [22]. In the NUMGEN project [23], a corpus of numerical expressions was collected and a formal model for planning specifications considering proportions was developed. The underlying theory and the design of the working program are described in [24].

There are different initiatives that propose guidelines that may help when rewriting a text to make it more comprehensive. Some of them are Plain Language [25], the European Guidelines for the Production of Easy-to-Read Information [26] and the Web Content Accessibility Guidelines [27].

III. SIMPLIFICATION OF NUMERICAL INFORMATION

The way in which information is written can exclude many people, especially those who have problems to read, write or understand. Nowadays, plenty of information in daily news or reports comes in the form of numerical expressions, for example, economic statistics, demography data, and in a special case, the measure of the ingredients of recipes.

A first possible approach to solve this important social problem is making numerical information accessible by rewriting difficult numerical expressions in a simpler way.

In this paper, we consider cookery recipes as a special case of study in which to analyse the different representations used for numerical information. We work with different kinds of Spanish recipes.

A. Recipes as a Special Case of Study

Analysing several kinds of recipes we can observe the ingredients of a recipe are often presented in very different ways. The main features observed are two: the units and the mathematical representation used.

Through out a recipe, some ingredients are shown in different units or different mathematical representations. Using different units of measure for the ingredients can be a problem for the people reading the recipe, and can lead to confusion. Explaining these quantities in different mathematical representations does not help either, as it can make it considerable more difficult to understand what quantity is meant exactly.

For example, the recipe for a pudding explains that “500 ml” of milk are required, and later indicates that “0.5 l (1/2 l)” milk should be poured in a saucepan. Different units “ml” and “l” (millilitres and litre) are used for the same ingredient. Additionally, the mathematical form also changes. The first time that the milk is mentioned a whole number is used “500” and the second time both a decimal number, “0.5”, and a fraction, “1/2 l”, are used to explain the meaning with different mathematical representations. From the analysis of this brief example, we propose the following assumptions:

- The units used for the ingredients influence the mathematical representation of the quantity.
- The mathematical form chosen is relevant to how well the meaning of the quantity is understood.

B. Testing our assumptions

We plan to carry out a user study in the domain of cooking, both with non-experts and experts in cooking. Non-experts’

preferences will represent the most understandable values at which we are aiming.

For the experiment, we need to select a set of different kinds of recipes -- appetizers, starters, desserts, etc. -- focusing on the numerical information in the ingredients of the recipe. We want to know if the units used and the mathematical representation of the quantity play a role in the level of comprehension of the recipe. Once we carry out a study based on a questionnaire, we can measure text comprehension of each recipe and for each participant.

The experiment will follow a within-subjects design, so every participant contributes to each of the conditions in the experiments. The order of conditions must be counter-balanced to cancel out sequence effects. Text comprehension is measured according to the correct answers from the questionnaires.

Multiple-choice questions will have three possible choices, one correct and two incorrect. From these answers, we will compute the percentage of correct answers, where the correct choice scored 100% and the others 0%. In addition, each participant must fill in a questionnaire for their preferences in terms of the numerical information representation. A Likert scale of the statements can be used to ask about how easy was to comprehend the recipe focusing on the quantities of the ingredients.

Participants will be asked to use their judgement and previous knowledge to decide what kind of representation of the numerical information for the ingredients is better understood.

With all this information we could develop an automatic system to rewrite difficult numerical expressions in recipes in a simpler way. For this we will follow the preferences of units and mathematical representations previously collected.

IV. DISCUSSION

The special case of study proposed in this paper, the task of cooking a given dish from a written recipe, is a task that many users may engage in within their everyday lives, with a significant presence of numerical information in textual form, and where this numerical information is very relevant to the outcome of the task. For the task to be carried out successfully by all, the accessibility of this numerical information needs to be as high as possible.

Another important feature is the language. As we are working with Spanish recipes, our analysis is based on the use of Spanish measures. However, for other languages, like English, there are different measures for different ingredients (oz, fl, etc). So, the language of the original recipe that we want to adapt is important to determinate the kind of units we can use in the adaptation.

The different units used are often difficult to understand because they come in formats or units that the user is not familiar with. For instance, many recipes use rare units, like “dl” (decilitre) for different ingredients, and this is difficult to understand and to compare with usual units. An adaptation process should choose a particular unit and a particular format and apply it consistently i.e. “1 dl” or “0.1 l” or “1/10 l”.

In this sense, a possible operation to improve understandability would be to adapt the units in more unusual formats to more frequently used ones. This work is based on the assumption that units and formats more frequently used are easier to understand.

Finally, previous work in text simplification has focused in numerical information. The conclusions of this work, as well as the simplification strategies used, should be considered in any attempt towards the implementation of an automatic simplification system.

V. CONCLUSIONS AND FUTURE WORK

The presence of numerical information in a text impacts its readability, as can be seen in previous work [28]. In this paper we have presented a preliminary analysis of the type of numerical information shown in recipes. This first step is previous to any attempt to automatize of the process, as it is required to know what kind of phenomena can be expected. We can conclude that the first step towards a useful formalization of how to adapt the numerical information in recipes has been taken.

The process of adaptation of texts by hand is extremely time and effort consuming so any attempts to automate part of this process can leverage the access to information. The final objective of this research is the development of an automatic system to adapt the different representations of quantities and units used in recipes and to ease the process of text adaptation for reading-impaired people. Depending on the final user we will study what kind of representations are needed and which their preferences are.

ACKNOWLEDGMENT

This research is partially funded by the Spanish Ministry of Education and Science (TIN2009-14659-C03-01 Project), and the FPI grant program.

REFERENCES

- [1] United Nations. Standard Rules on the Equalization of Opportunities for Persons with Disabilities. Technical Report. 1994.
- [2] S. Bautista, R. Hervás, P. Gervás, R. Power, S. Williams. How to Make Numerical Information Accessible: Experimental Identification of Simplification Strategies. In Proc. of the 13th IFIP TC13 Conference on Human-Computer Interaction. Lisbon, Portugal. 2011.
- [3] S. Bautista, R. Hervás, P. Gervás, R. Power, S. Williams. Experimental Identification of the Use of Hedges in the Simplification of Numerical Expressions. In Proc. of the Second Workshop on Speech and Language Processing for Assistive Technologies. Edinburgh, Scotland. 2011.
- [4] S. Bautista, R. Hervás, P. Gervás, R. Power, S. Williams. A System for the Simplification of Numerical Expressions at Different Levels of Understandability. In Proc. of the Workshop Natural Language Processing for Improving Textual Accessibility. Atlanta, USA.
- [5] R. Chandrasekar, C. Doran, B. Srinivas. Motivations and methods for text simplification. In Proc. of the Sixteenth International Conference on Computational Linguistics, pages 1041–1044. 1996.
- [6] A. Siddharthan. Syntactic Simplification and Text Cohesion. Ph.D. thesis, University of Cambridge. 2003.
- [7] A. Candido Jr., E. Maziero, C. Gasperin, T. A. S. Pardo, L. Specia, S. M. Aluisio. Supporting the Adaptation of Texts for Poor Literacy Readers: a Text Simplification Editor for Brazilian Portuguese. In Proc. of the Workshop on Innovative Use of NLP for Building Educational Applications, pages 34–42, 2009.
- [8] S. Bautista, P. Gervás, I. Madrid. Feasibility Analysis for SemiAutomatic Conversion of Text to Improve Readability. In Proc. of The Second International Conference on Information and Communication Technologies and Accessibility, Hammamet, Tunisia, May. 2009.
- [9] W. Daelemans, A. Hothker, E. Tjong Kim Sang. Automatic Sentence Simplification for Subtitling in Dutch and English. In Proc. of the 4th Conference on Language Resources and Evaluation, pages 1045–1048, Lisbon, Portugal. 2004.
- [10] S. E. Petersen and M. Ostendorf. Text Simplification for Language Learners: A Corpus Analysis. In Proc. of Workshop on Speech and Language Technology for Education. 2007.
- [11] C. Gasperin, L. Specia, T. F. Pereira, S. M. Aluisio. Learning when to simplify sentences for natural text simplification. In Proc. of the Encontro Nacional de Inteligencia Artificial, pages 809–818, Bento Gonalves, Brazil. 2009.
- [12] S. Devlin, J. Tait. The use of a Psycholinguistic database in the Simplification of Text for Aphasic Readers. Lecture Notes. Stanford, USA: CSLA. 1998.
- [13] Y. Canning. Cohesive simplification of newspaper text for aphasic readers. In 3rd annual CLUK Doctoral Research Colloquium. 2000.
- [14] R. Chandrasekar, B. Srinivas. Automatic induction of rules for text simplification. Knowledge-Based Systems, 10. 1997.
- [15] A. Siddharthan. Resolving attachment and clause boundary ambiguities for simplifying relative clause constructs. In Proc. of the Student Research Workshop, 40th Meeting of the ACL. 2002.
- [16] S. Williams, E. Reiter, L. Osman. Experiments with discourse-level choices and readability. In Proc. of the European Natural Language Generation Workshop and 11th Conference of the EACL, pages 127–134. 2003.
- [17] I. H. Khan, K. Deemter, G. Ritchie. Generation of referring expressions: managing structural ambiguities. In Proc. of the 22nd International Conference on Computational Linguistics, pages 433–440, Manchester. 2008.
- [18] S. Williams, E. Reiter. Generating readable texts for readers with low basic skills. In Proc. of the 10th European Workshop on Natural Language Generation, pages 140–147, Aberdeen, Scotland. 2005.
- [19] E. Peters, J. Hibbard, P. Slovic, N. Dieckmann. Numeracy skill and the communication, comprehension, and use of risk-benefit information. Health Affairs, 26(3):741–748. 2007.
- [20] N. Dieckmann, P. Slovic, E. Peters. The use of narrative evidence and explicit likelihood by decision makers varying in numeracy. Risk Analysis, 29(10). 2009.
- [21] A. M. Bisantz, S. Schinzing, J. Munch. Displaying uncertainty: Investigating the effects of display format and specificity. Human Factors: The Journal of the Human Factors and Ergonomics Society, 47(4):777. 2005.
- [22] H. Mishra, A. Mishra, B. Shiv. In praise of vagueness: malleability of vague information as a performance booster. Psychological Science, 22(6):733–8, April. 2011.
- [23] S. Williams, R. Power. Precision and mathematical form in first and subsequent mentions of numerical facts and their relation to document structure. In Proc. of the 12th European Workshop on Natural Language Generation, Athens. 2009.
- [24] R. Power and S. Williams. Generating numerical approximations. Computational Linguistics, 38(1). 2012.
- [25] (2005) The Plain Language Action and Information Network (PLAIN) [Online]. Available: <http://www.plainlanguage.gov/>
- [26] (1998) The European Guidelines for the production of easy-to-read information. [Online]. Available: <http://www.disabilityrightsfund.org/es/node/510>
- [27] (2008) W3C Web Content Accessibility guidelines. [Online]. Available: <http://www.w3.org/TR/WCAG20/>
- [28] L. Rello, S. Bautista, R. Baeza-Yates, P. Gervás, R. Hervás, H. Saggion. One Half or 50%? An Eye-Tracking Study of Number Representation Readability. In Proc. of the 14th IFIP TC13 Conference on Human-Computer Interaction. Cape Town, South Africa. 2013.