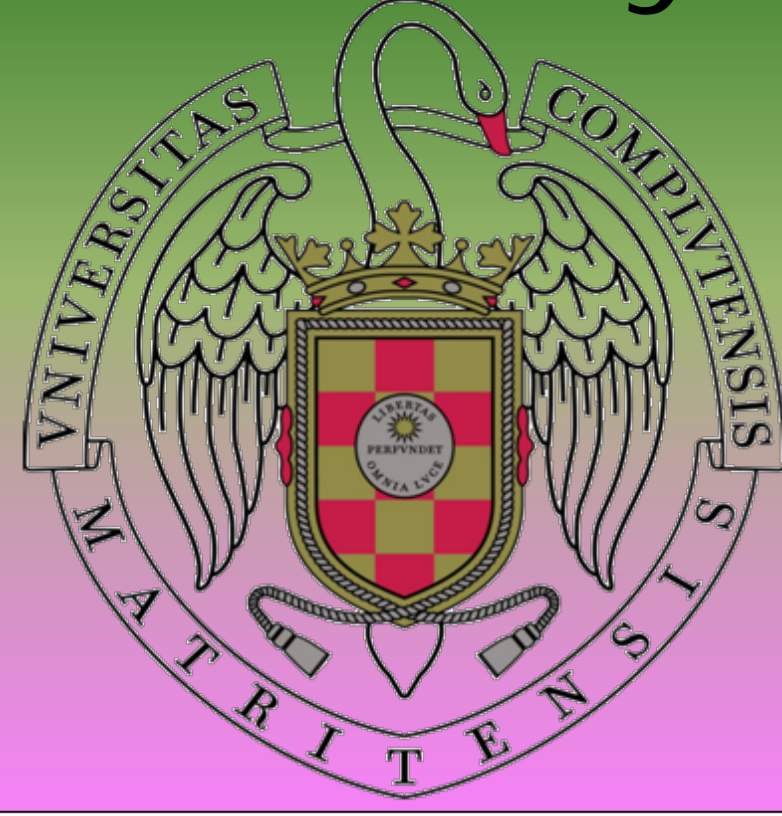# A System for the Simplification of Numerical Expressions at Different Levels of Understandability

Susana Bautista, Raquel Hervás, Pablo Gervás
Universidad Complutense de Madrid, Spain
{subautis,raquelhb} @fdi.ucm.es, pgervas@sip.ucm.es

Richard Power, Sandra Williams
Department of Computing, The OU, Milton Keynes , UK
{r.power, s.h.williams} @open.ac.uk

The Open University

ABSTRACT

Numerical expressions convey different information in texts, but many times this kind of information is difficult to understand for many people, including non-native speakers, and people with disabilities. The purpose of this paper is to motivate and describe a system that simplifies numerical expressions in English texts, along with an evaluation study in which experts in numeracy and literacy assessed the outputs of this system. We have worked with a collection of newspaper articles with a significant number of numerical expressions. The results are discussed in comparison to conclusions obtained from a prior empirical survey.

NUMERICAL EXPRESSION

About   20%
Hedge   Quantity

Level Difficulty

- *Fractions*
- *Dec. Percent.*
- *Percent.*

Obtain List of Candidates

```
<sentence id="7">
   Overall figures showed the national
pass rate soared
<numex
   hedge= "more than"
   hedge-sem="greaterthan"
   type="percentages"
   format="digits"
   Vg="0.28 "
>
   more than 28%
</numex>
...
```

Proportion= 0.28
P=28%
P~30%
P< 30%
F~1/3
F<1/3
P~33%
P< 33%
F~1/4
F>1/4
P~25%
P>25%

Select Candidate

F~1/3

Hedge Choice

Original NumExp	If Candidate > Original	If Candidate = Original	If Canidate < Original
More than OrigValue	*Around Candidate*	*More than Candidate*	*More than Candidate*
Exactly OrigValue	*Less than Candidate*	*Exactly Candidate*	*More than Candidate*
Less than OrigValue	*Less than Candidate*	*Less than Candidate*	*Around Candidate*
OrigValue	*Around Candidate*	*Candidate*	*Around Candidate*

Candidate > Original
0.33>0.28
→ Around

SYSTEM OPERATION

Output Expression

Simplied Expression

It is **around 1/3**

Parser

Original Expression

It is **more than 28%**

STARTING HYPOTHESES

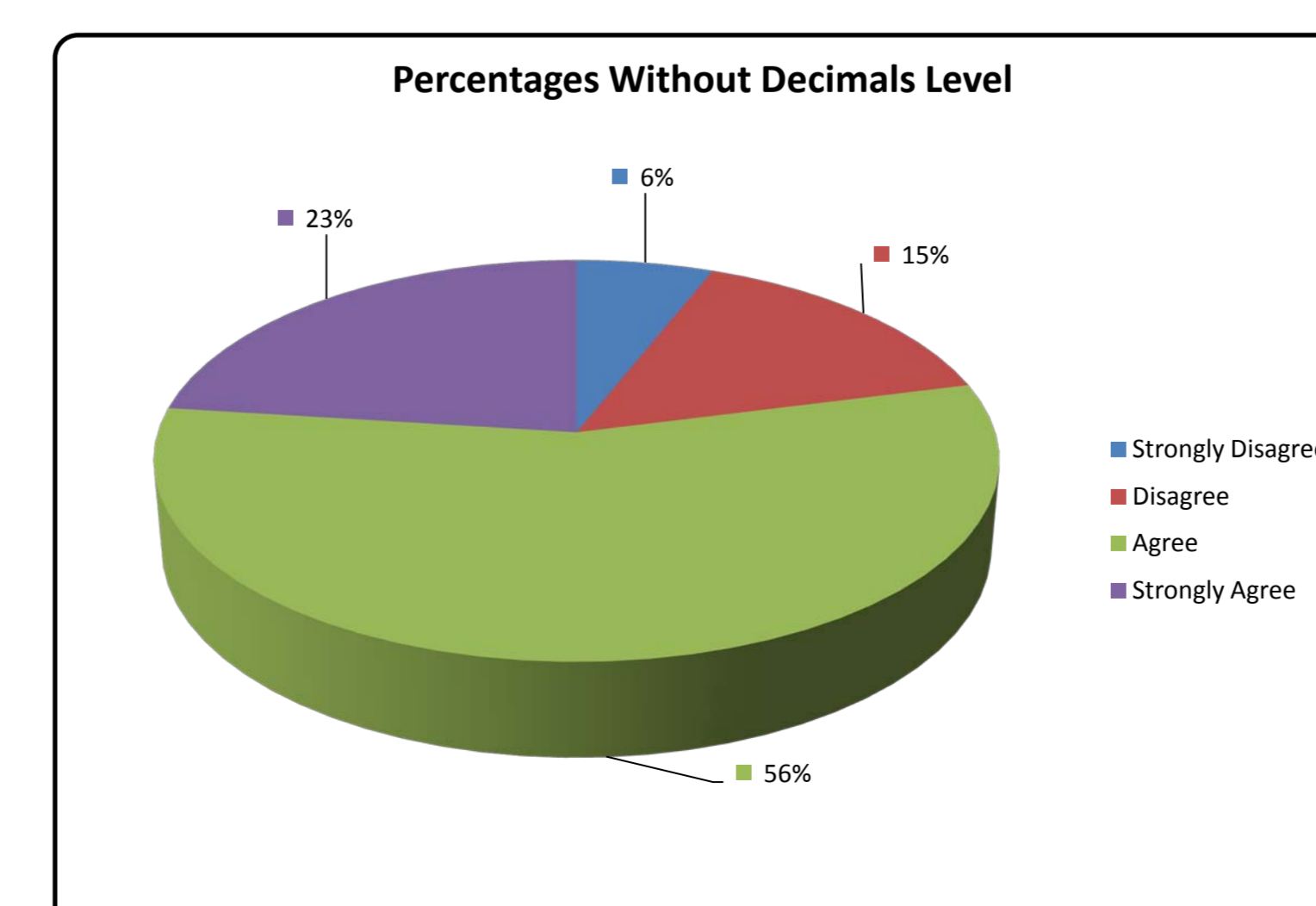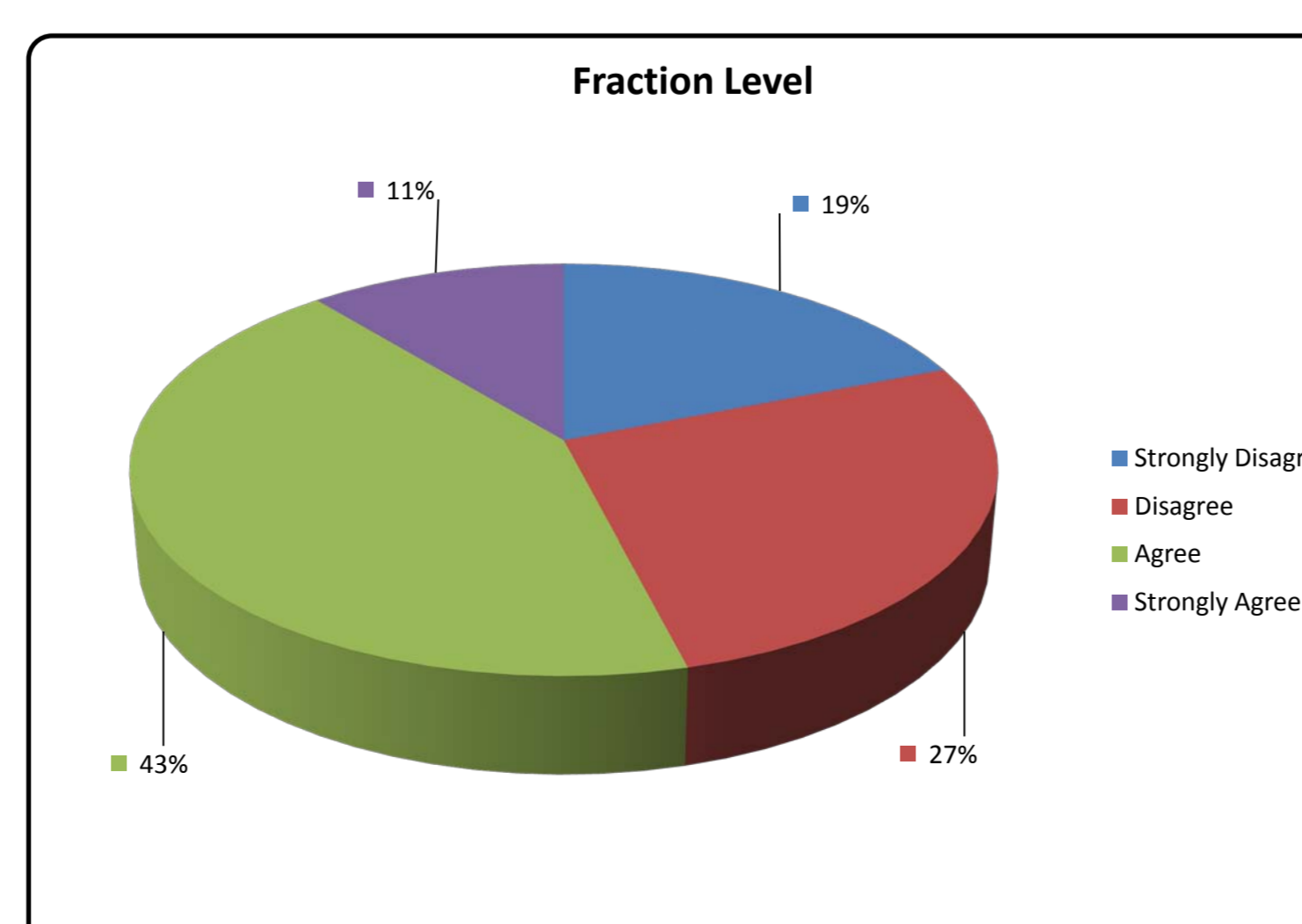| Round and common values are prefered | 29.3%→ almost 30% |
	74.8%→ just under 3/4
Value of the original expression influences the choice of the simplification strategies	*Extreme values*
	97%→ almost all
	0.8%→ almost none
	Central values
	48%→ around 50%
Hedges are used only if there is loss of precision	*No Hedges*
	25%→ a quarter
	Hedge
	53%→ just over a half

SURVEY

http://www.surveymonkey.com/s/WJ69L86
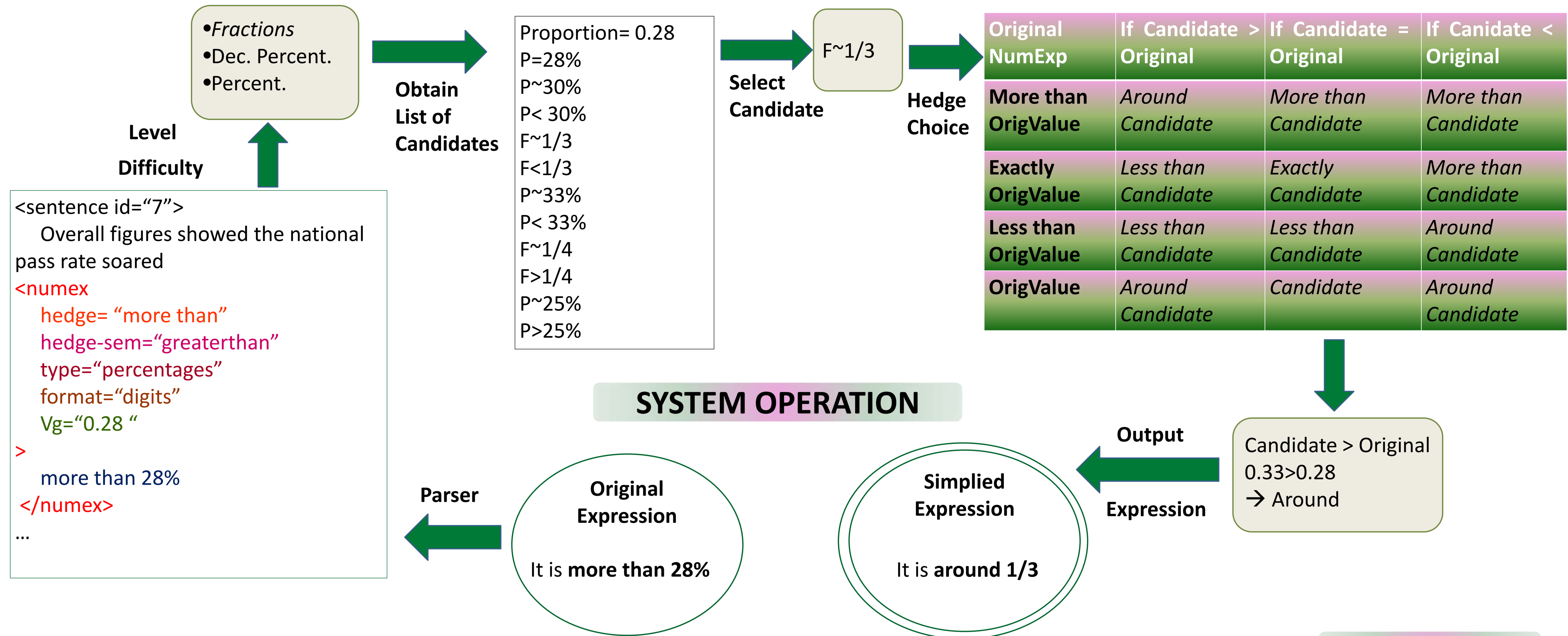
32. A: The proportion of exams awarded B grades or better rose to [50.8 per cent] this year, and [73.9 per cent] scored at least a C.

	Strongly disagree	Disagree	Agree	Strongly agree
B: The proportion of exams awarded B grades or better rose to [around 1/2] this year, and [around 3/4] scored at least a C.	○	○	○	○
C: The proportion of exams awarded B grades or better rose to [around 51%] this year, and [around 74%] scored at least a C.	○	○	○	○

Optional: add a comment

Fraction Level
19%
11%
43%
27%
Strongly Disagree
Disagree
Agree
Strongly Agree

Percentages Without Decimals Level
6%
15%
23%
56%
Strongly Disagree
Disagree
Agree
Strongly Agree

CONCLUSIONS

- ▶Definition of a scale of difficulty of numerical expressions.

- ▶Identification of the rules for:
  - ▶ the seleccion of candidate substitution
  - ▶the application of hedges

- ▶Evaluation with experts results in acceptable rates of agreement.

natural interaction based on language

http://nil.fdi.ucm.es/   http://mcs.open.ac.uk/nlg/