FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

EURO
THE ASSOCIATION OF EUROPEAN OPERATIONAL RESEARCH SOCIETIES

Investigating Parallel Implementaions of Genetic Algorithms for Stochastic Part-of-Speech Tagging

Shimanto Rahman, prof. dr. Matthias Bogaert & prof. dr. Dirk Van den Poel

GHENT UNIVERSITY

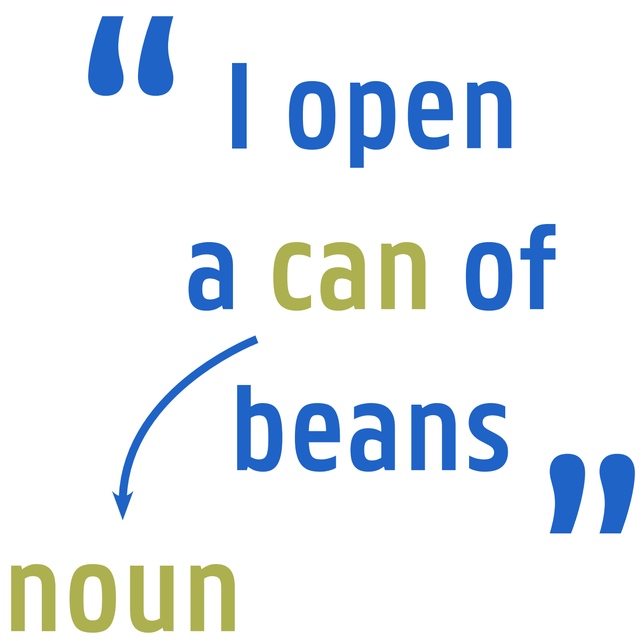
What I will be talking about

01 Problem statement
What is Part-of-Speech tagging? Why is it useful?

02 Methods
Which methods are used? How are they applied?

03 Results
What are the main takeaways?

Words are ambigious



GHENT UNIVERSITY

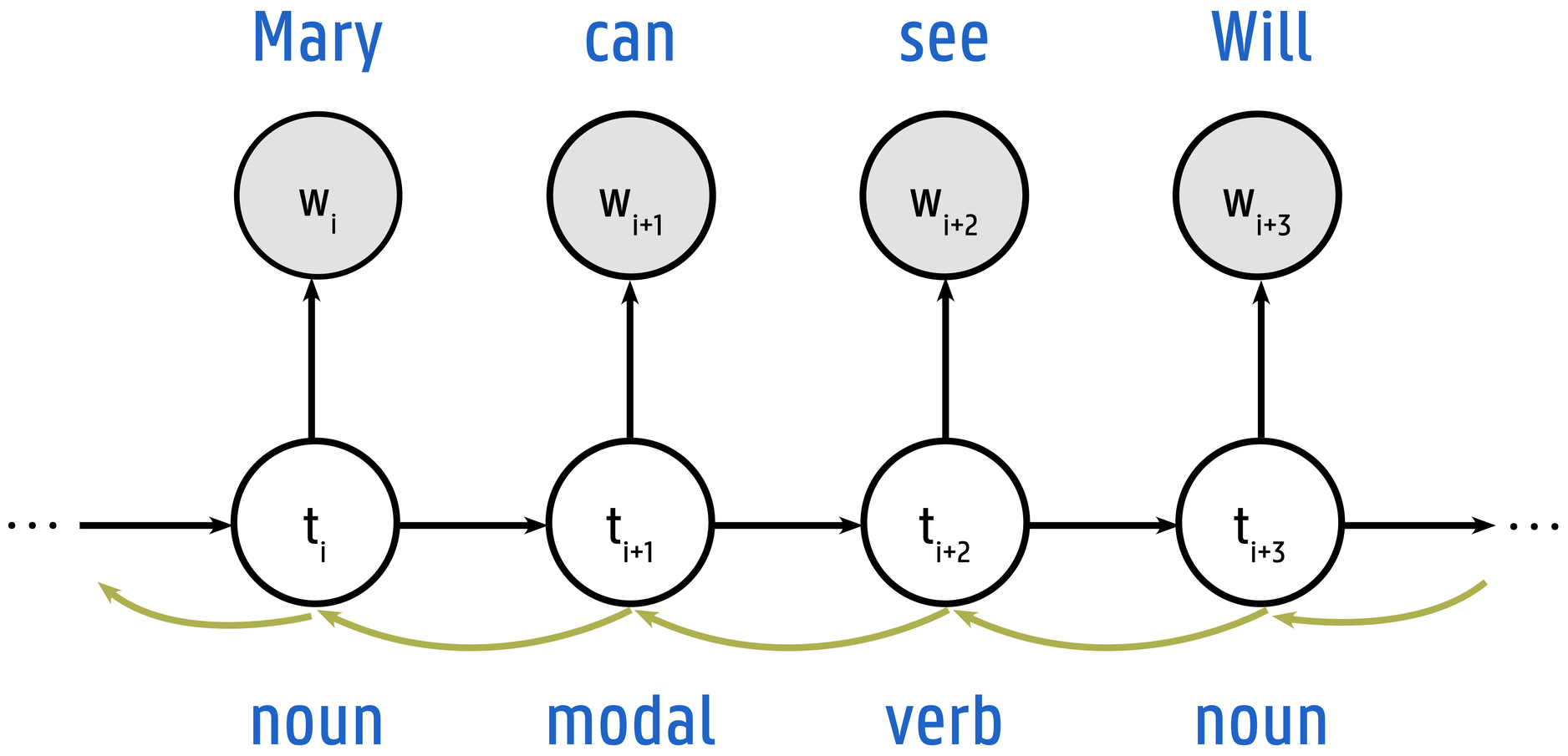
What I will be talking about

01 **Problem statement**
What is Part-of-Speech tagging? Why is it useful?

02 **Methods**
Which methods are used? How are they applied?

03 **Results**
What are the main takeaways?

Hidden Markov Model



GHENT UNIVERSITY

Literature Review

Publication	GA	Distributed GA	Cellular GA
Aroujo (2002)	X		
Aroujo (2004)	X		
Aroujo, Luque & Alba (2004)	X	X	
Aroujo, Luque & Alba (2006)	X	X	
This study	X	X	X

GHENT UNIVERSITY

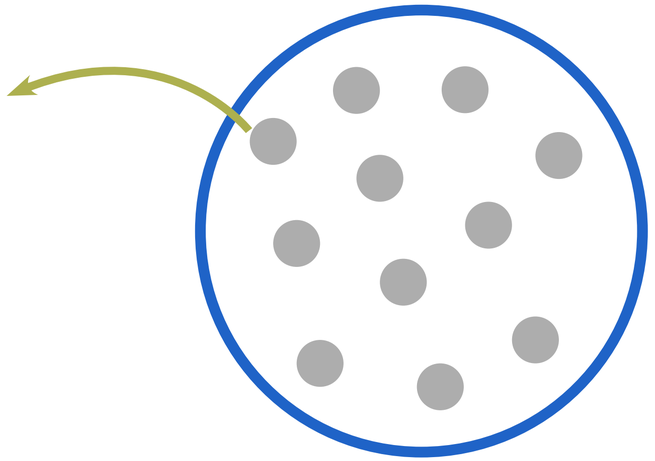
Literature Review

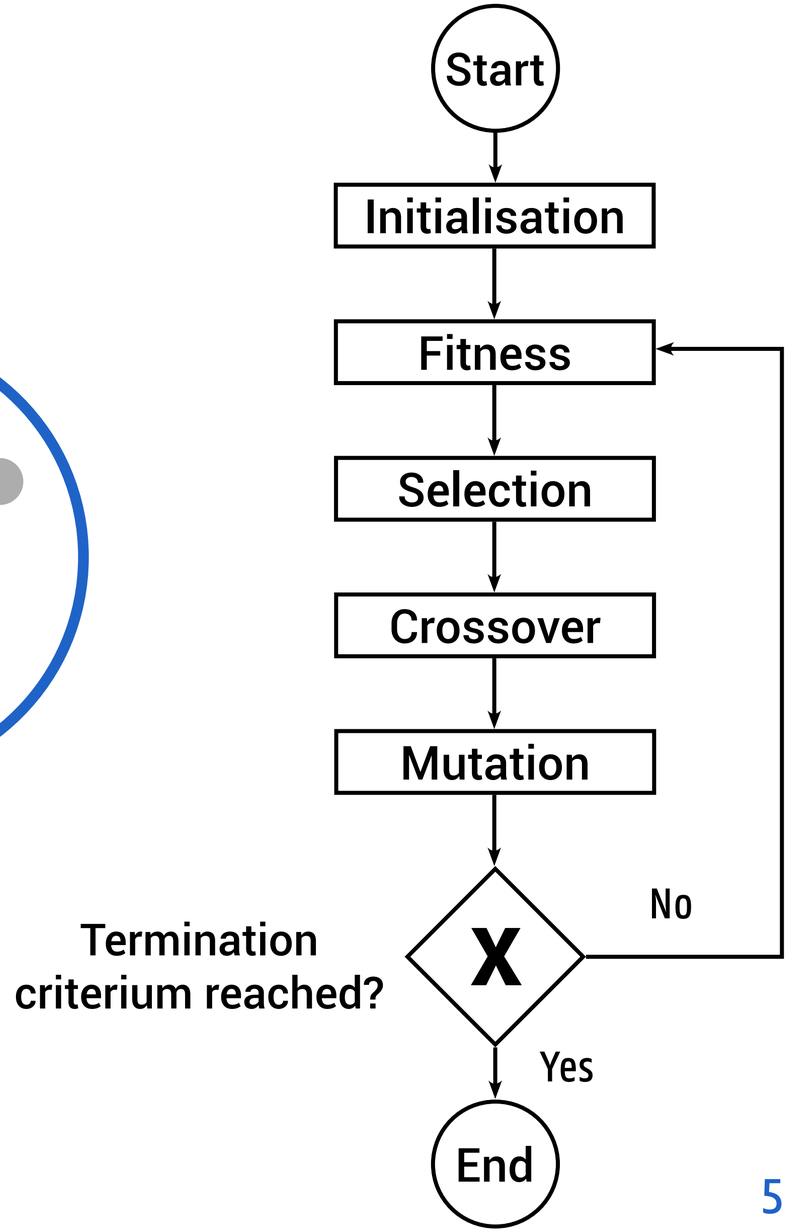
Publication	Token Accuracy	Sentence Accuracy	Unknown Token Accuracy
Aroujo (2002)	X		
Aroujo (2004)	X		
Aroujo, Luque & Alba (2004)	X		
Aroujo, Luque & Alba (2006)	X		
This study	X	X	X

GHENT UNIVERSITY

Genetic Algorithm

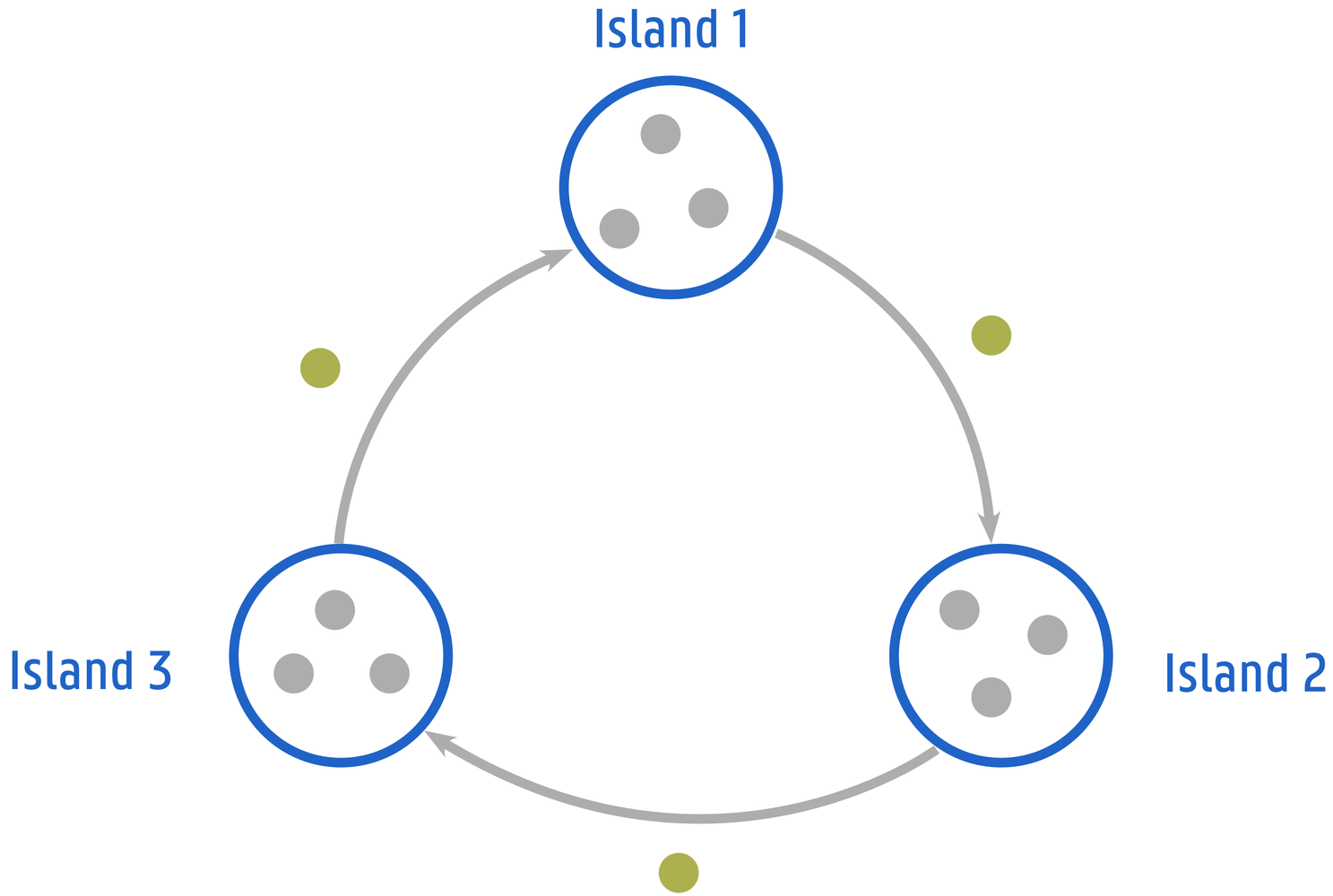
Tokens	Mary	can	see	Will
Phenotype	NOUN	MODAL	VERB	NOUN
Genotype	0	1	2	0





GHENT UNIVERSITY

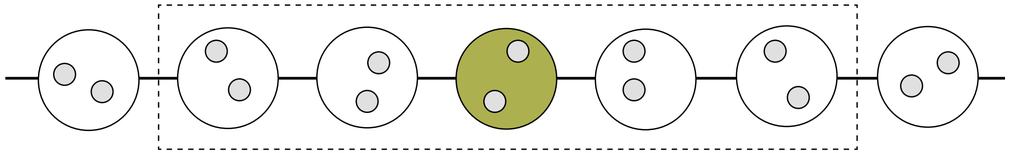
Distributed Genetic Algorithm



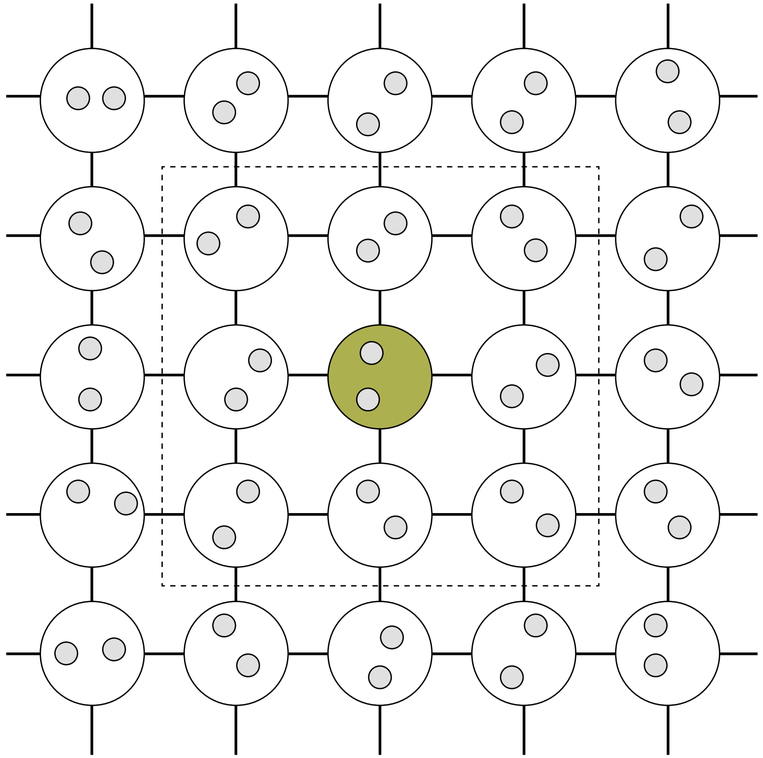
GHENT
UNIVERSITY

Cellular Genetic Algorithm

Circular



Toroidal



GHENT UNIVERSITY

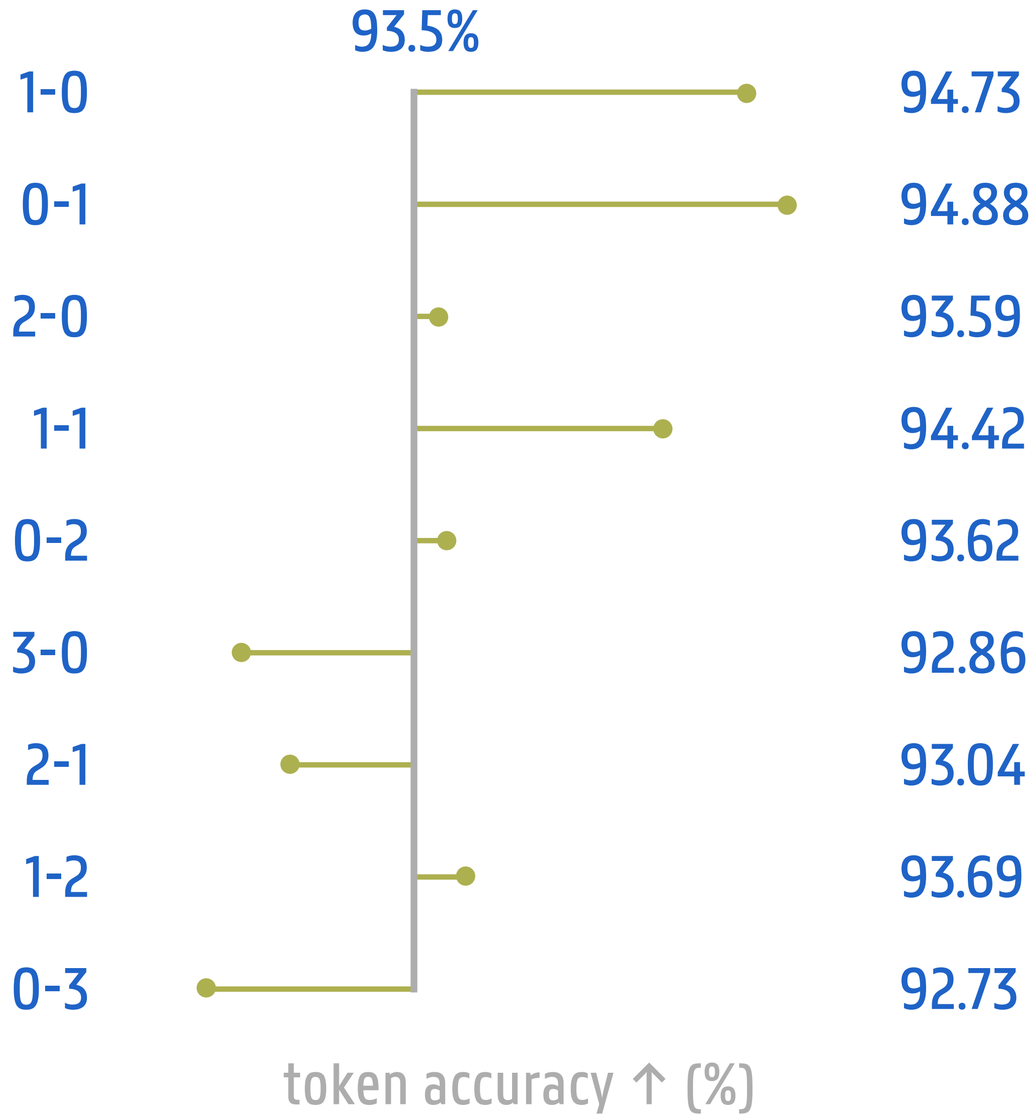
What I will be talking about

01 **Problem statement**
What is Part-of-Speech tagging? Why is it useful?

02 **Methods**
Which methods are used? How are they applied?

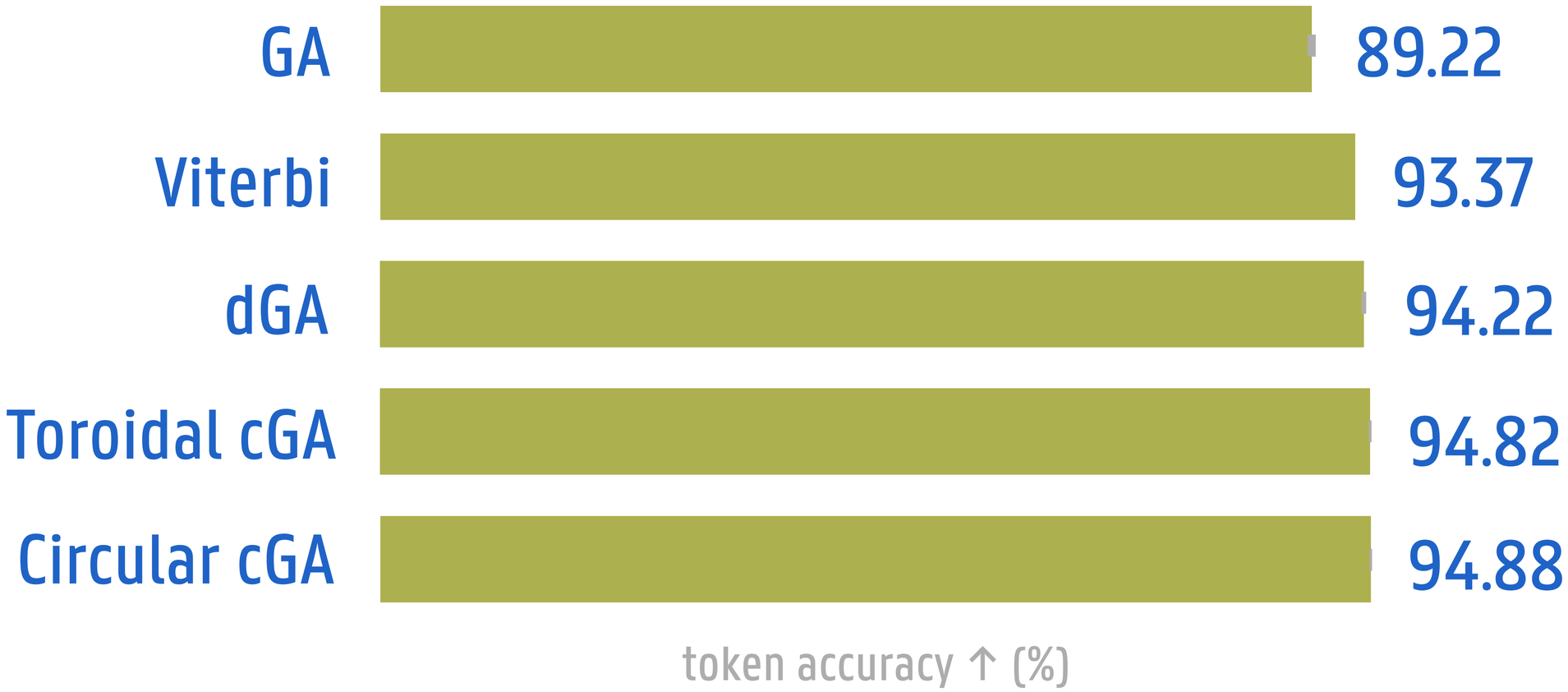
03 **Results**
What are the main takeaways?

Smaller context sizes perform best



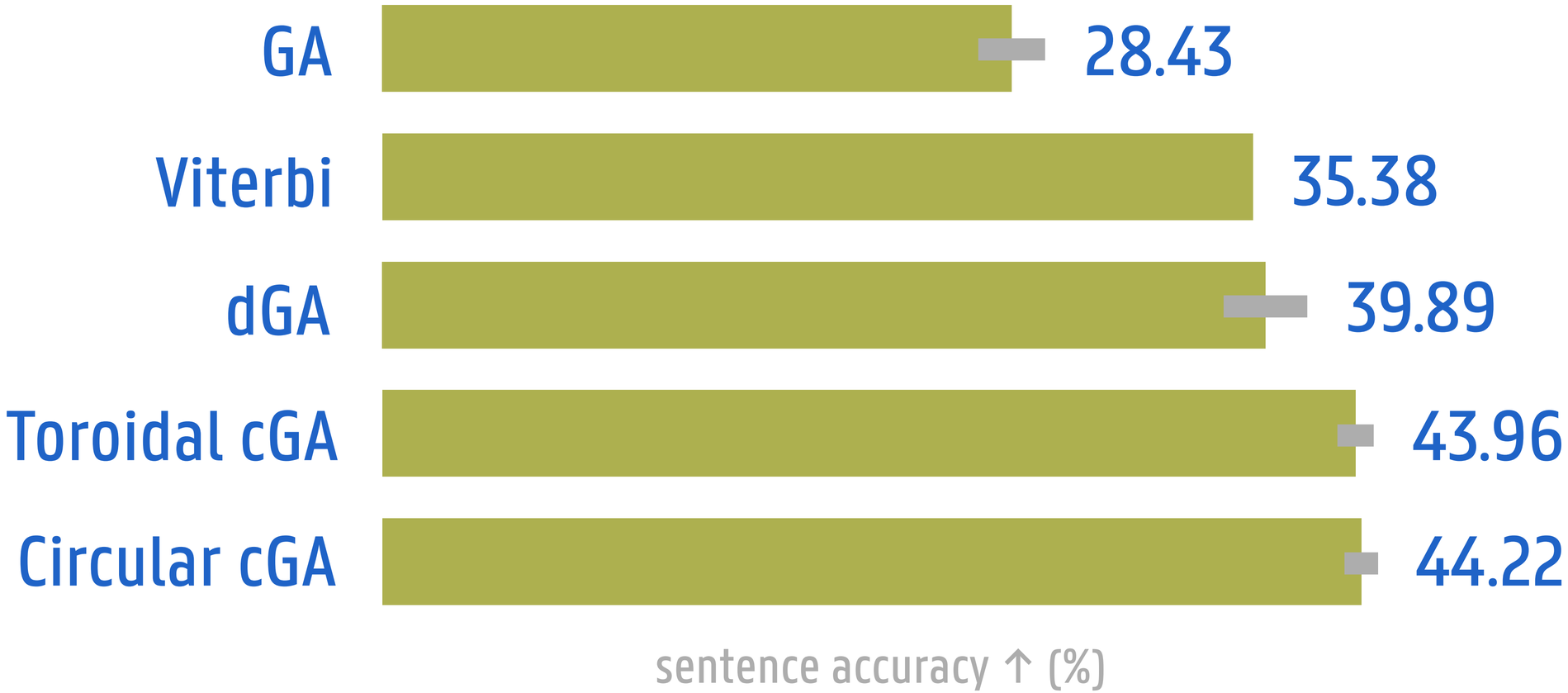
GHENT UNIVERSITY

Traditional GA prematurely converges



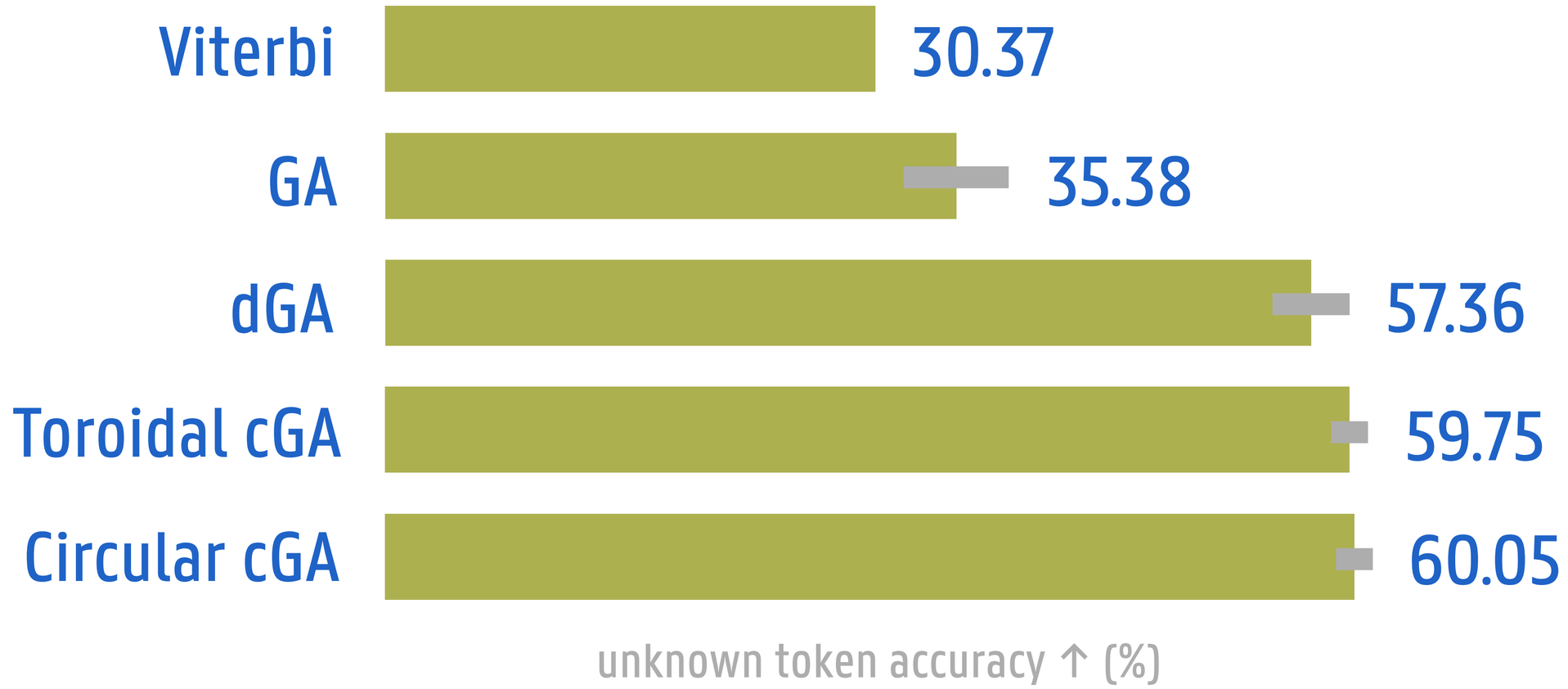
GHENT UNIVERSITY

Correctly classifying a sentence remains hard



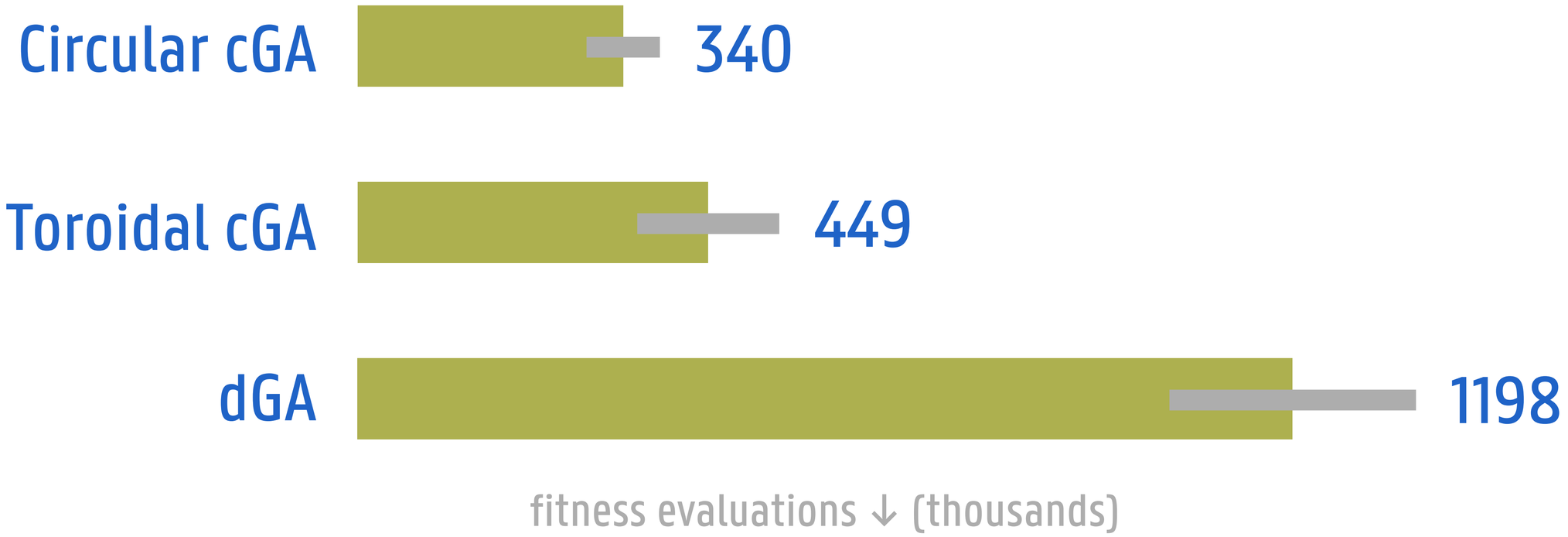
GHENT UNIVERSITY

Parallel GAs are more robust against unknown words



GHENT UNIVERSITY

Circular topology converges faster



GHENT UNIVERSITY

Conclusions

1

cGAs perform best

across all performance measures over traditional GAs and dGAs

2

Circular topology

converges faster to a local optimum than a toroidal topology

3

Smaller context sizes

are preferred over larger context sizes for this simple tag set

GHENT UNIVERSITY

Further Research

1

More complex tag sets

may have an effect on the impact on larger context sizes

2

More diverse corpora

will make the generalizability of the study more robust

GHENT UNIVERSITY

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

EURO
THE ASSOCIATION OF EUROPEAN OPERATIONAL RESEARCH SOCIETIES

www.github.com/ShimantoRahman/EURO2022

www.linkedin.com/in/shimanto-rahman

shimanto.rahman@ugent.be

GHENT UNIVERSITY