































corpus. We plan to find a larger corpus to increase the size of the training corpus and test the effectiveness of the parameters we have defined for tagging the Twi language.

## **8. Conclusion**

In this study, we presented an approach to tagging the parts of speech of Twi. We worked with a corpus of 36,000 words. We used a tagger (TreeTagger) to test its effectiveness by analysing the results obtained. The tagging took place over three training phases, at the end of which the automatic tagging of the entire corpus was done. Our work has contributed to the enrichment of studies that focus on Twi since we shed light on fundamental phenomena that have been the subject of previous studies. Also, this work is a first step in the production of a database that will be integrated into linguistic tools for the automatic processing of the Twi language. This database is composed of lexical words endowed with morphological information by means of the automatic tagging of the Twi that we carried out within the framework of this work.

## **Acknowledgements**

We would like to thank Mr. Scannell who provided us with the corpus on the Twi language. This corpus formed the basis of our work. We also thank all those who helped us financially.

## References

- [1] Adu Manyah, K. (2002) *Introduction à la phonétique et à la phonologie africaines: les sons de tous les jours*. Paris, France : L'Harmattan.
- [2] Agyekum, K. (2010) *Akan kasa nhyehyee*. Accra : Dwumfuor Publications.
- [3] Brill, E. (1992) A simple rule-based part of speech tagger. *Proceedings of the workshop on Speech and Natural Language*, 112–116. <http://dl.acm.org/citation.cfm?id=1075553>
- [4] Christaller, J. G., (1964) A grammar of the Asante and Fante language called Tshi Chwee, Twi : based on the Akuapem dialect with reference to the other Akan and Fante dialects repr. of 1875 . Farnborough, Hants.: Gregg Press.
- [5] Ghahramani, Z. (2001) An Introduction to Hidden Markov Model and Bayesian Networks. *International Journal of Pattern Recognition and Artificial Intelligence*. 15(1), 9-42.
- [6] Greenberg, J. H. (1963) *The Languages of Africa*. *International journal of American linguistics*, 29(1).
- [7] Hasan, F. M., Uzzaman, N., & Khan, M. (2007) Comparison of different POS Tagging Techniques (N-Gram, HMM and Brill's tagger) for Bangla. *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, 121–126. [http://link.springer.com/chapter/10.1007/978-1-4020-6264-3\\_23](http://link.springer.com/chapter/10.1007/978-1-4020-6264-3_23)
- [8] Helmut S. (1994) [Probabilistic Part-of-Speech Tagging Using Decision Trees](http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf). *Proceedings of International Conference on New Methods in Language Processing*. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>
- [9] Helmut S. (1995) [Improvements in Part-of-Speech Tagging with an Application to German](http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger2.pdf). *Proceedings of the ACL SIGDAT-Workshop*. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger2.pdf>
- [10] Humera Khanam, M., Madhumurthy, K.V. & Khudhus, M.D.A. (2013) Comparison of TnT, Max. Ent, CRF Taggers for Urdu Language. *International Journal of Engineering Sciences Research-IJESR*. 4(1), 1164-1168.
- [11] Sawalha, M., & Atwell, E. S. (2010) Fine-grain morphological analyzer and part-of-speech tagger for Arabic text. *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)* 1258–1265. <http://eprints.whiterose.ac.uk/42641/>
- [12] Scannell, K. P. (2007) The Crúbadán Project: Corpus building for under resourced languages. Building and Exploring Web Corpora. *Proceedings of the 3<sup>rd</sup> Web as Corpus Workshop*, 5–15.



## Annexe 1

DEF	Definite Article
INDEF	Indefinite Article
NAbs	Abstract Nom
PRel	Relative Pronoun
PrDEM	Demonstrative Pronoun
PrINDEF	Indefinite Pronoun
PrPers	Personal Pronoun
Pin	Interrogative Pronoun
P	Pronoun
Nc	Common Noun
N	Noun
NP	Proper Nom
PC	Past Tense
IMP	Imperfect
Imp	Imperfect Aspect
Perf	Perfect Aspect
FUTP	Near Future
FUTL	Distant Future
ADJ-DEM	Demonstrative Adjective
ADJPO	Possessive Adjective
ADJ-Int	Interrogative Adjective
AQA	Qualitative Adjective
AQE	Qualitative Adjective
Pr-PO	Possessive Pronoun
EMP	Emphasis Marker
NG	Negation
PAT	Particle
ME	Emphasis Marker
AdvL	Adverb of place
AdvT	Adverb of time
AdvM	Adverb of Manner
AdvD	Adverb of Degree

GSJ

AdvF	Adverb of Frequency
ANC	Cardinal Numeral Adjective
Card	Cardinal Adjective
ANO	Ordinal Adjective
ConCor	Conjunction of coordination
ConSub	Conjunction of subordination
Pper	Personal Pronoun
1PS	First person singular
2PS	Second person singular
3PS	Third person singular
1PP	First person plural
2PP	Second person plural
3PP	Third person plural
V	Verb
Imper	Imperative
Int	Interjection
Con	Connector
Prog	Progressive
Pr	Present
Sg	Singular
Pl	Plural



## Annexe 2

Token	TreeTag	Good Tag
w'ayi	Nc	
wɔn	3PP	
Adi	PerfV	
pɛn	Nc	EMP
No	DEF	
Ahyɛase	Nc	
Foforo	Nc	Adv
;	SENT	
Ne	Concor	
yɛn	1PP	
Botae	Nc	
sɛ	PRel	
yɛbɛboa	Nc	1PPFUTV
Sukuufo	Nc	
Ama	PerfV	
wɔn	3PP	
Aduru	Nc	
wɔn	3PP	
Botae	Nc	
Ho	POST	
.	SENT	
Woa	Nc	2PS
hwɛ	Nc	V
Family	Nc	
Guy	Nc	
,	SENT	
dieɛ	Nc	
Brian	Nc	
De	V	
Stewie	Nc	
Ka	Nc	
No	DEF	
?	SENT	
Sɛ	PRel	
yɛyɛ	Nc	1PPV
Obi	Nc	PrDem
bɔne	Nc	Adj
A	EMP	
,	SENT	
yɛsrɛ	Nc	1PPV
bɔne	Nc	Adj
fakye	Nc	VV

© GSJ

<b>Na</b>	Concor
<b>Ama</b>	PerfV
<b>yene</b>	Nc
<b>Onii</b>	Nc PrDem
<b>korə</b>	Nc
<b>No</b>	DEF
<b>Ntam</b>	Nc
<b>Adwo</b>	Nc PerfV
<b>.</b>	SENT
<b>(</b>	SENT
<b>1</b>	Card
<b>)</b>	SENT
<b>Den</b>	Pin
<b>Nti</b>	EMP
<b>Na</b>	Concor
<b>wəfrɛ</b>	Nc
<b>No</b>	DEF
<b>“Asem</b>	Nc
<b>no”</b>	Nc
<b>?</b>	SENT
<b>Sɛ</b>	PRel
<b>wopɛ</b>	2PSV
<b>sɛ</b>	PRel
<b>Wo</b>	2PS
<b>De</b>	V
<b>Wo</b>	2PS
<b>Ho</b>	POST
<b>bə</b>	Nc V
<b>deɛ</b>	Pin
<b>Aka</b>	Nc
<b>No</b>	DEF
<b>A</b>	EMP
<b>,</b>	SENT
<b>ɛyɛ</b>	3PSV
<b>yɛ</b>	V
<b>,</b>	SENT
<b>ɛyɛ</b>	3PSV
<b>deɛ</b>	Pin
<b>Wo</b>	2PS
<b>pɛ</b>	V
<b>.</b>	SENT
<b>Baanodifoə</b>	Nc
<b>nsusuyɛ</b>	Nc
<b>Baanodifoə</b>	Nc
<b>De</b>	V
<b>wən</b>	3PP
<b>nsusuyɛ</b>	Nc



<b>A</b>	EMP	
<b>ɛfa</b>	3PSV	
<b>Sukuu</b>	Nc	
<b>No</b>	DEF	
<b>Ho</b>	POST	
<b>bɛma</b>	Nc	FUTV
<b>Sukuu</b>	Nc	
<b>Panyin</b>	Nc	
<b>.</b>	SENT	
<b>2</b>	Card	
<b>.</b>	SENT	
<b>Obiara</b>	Nc	

© GSJ