paṭi (படி) in Tamil:
A Corpus-based Investigation

THENNARASU SAKKAN
Central University of Kerala, Kerela

L. R. PREM KUMAR
LDC-IL, CIIL Mysore

AMUDHA R
LDC-IL, CIIL Mysore

ABSTRACT

One of the most frequently occurring words in Tamil corpus is paṭi *(படி), which shows various morpho-syntactic functions. Morphologically, the word can be analyzed either as a verb or a noun. In addition, it can become a particle when occurring after a noun, pronoun and relative participle, verbal participle. Given these categorical variations, the determining status of this lexical item* paṭi *has become a challenging endeavor in the field of NLP tasks such as POS tagging. In this way, we explore the various functional characteristics of* paṭi *and as a result, identify basic morpho-syntactic categories. With these descriptive accounts in view, we come to suggest for dealing with* paṭi *within the computational process like POS tagging and other advanced NLP related tasks. Finally, this paper lightly travels around how the artificial intelligence (AI) can be made to recognize the lexical* paṭi *by giving algorithms for resolving categorical ambiguities without recourse to further semantic level information. The output of human translation in various stirring environments is applied and fitted to the feasible algorithm, a recursive operation, which gives command to the NLP tools.*