

การวิเคราะห์ประโยคภาษาไทยตามแนวไวยากรณ์ฝั่งพา  
เพื่อใช้ในระบบการแจกลำประโยคด้วยคอมพิวเตอร์



นายวิโรจน์ อรุณมานะกุล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาอักษรศาสตรมหาบัณฑิต  
ภาควิชาภาษาศาสตร์

บัณฑิตวิทยาลัย จฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2533

ISBN 974-577-693-9

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จฬาลงกรณ์มหาวิทยาลัย

016416

10306390



A DEPENDENCY ANALYSIS OF THAI SENTENCES  
FOR A COMPUTERIZED PARSING SYSTEM

Mr. Wirote Aroonmanakun

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Arts

Department of Linguistics

Graduate School

Chulalongkorn University

1990

ISBN 974-577-693-9



Thesis Title : A Dependency Analysis of Thai Sentences  
for a Computerized Parsing System  
By : Mr. Wirote Aroonmanakun  
Department : Linguistics  
Thesis Advisor : Assistant Professor Peansiri Vongvipanond, Ph.D.  
Thesis Co-Advisor : Associate Professor Somchai Tayanyong

---

Accepted by the Graduate School, Chulalongkorn University in  
Partial Fulfillment of the Requirements for the Master's Degree.

*Thavorn Vajrabhaya*  
..... Dean of Graduate School  
(Professor Thavorn Vajrabhaya, Ph.D.)

Thesis Committee

*Pranee Kullavanijaya* Chairman  
(Associate Professor Pranee Kullavanijaya, Ph.D.)

*Peansiri Vongvipanond* Thesis Advisor  
(Assistant Professor Peansiri Vongvipanond, Ph.D.)

*S. Tayanyong*  
..... Thesis Co-Advisor  
(Associate Professor Somchai Tayanyong)

*Kingkarn Thepkanjana* Member  
(Kingkarn Thepkanjana, Ph.D.)

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว



วิโรจน์ อรุณมานะกุล : การวิเคราะห์ประโยคภาษาไทยตามแนวไวยากรณ์ฟังก์พ เพื่อใช้ในระบบ  
การแจงส่วนประโยคด้วยคอมพิวเตอร์ (A DEPENDENCY ANALYSIS OF THAI SENTENCES  
FOR A COMPUTERIZED PARSING SYSTEM) อ.ที่ปรึกษา : ผศ.ดร.เพ็ญศิริ วงศ์วิภานนท์,  
อ.ที่ปรึกษาร่วม : รศ.สมชาย ทยานยง, 242 หน้า. ISBN 974-577-693-9

งานวิจัยนี้มีวัตถุประสงค์ที่จะวิเคราะห์ประโยคภาษาไทยตามแนวทฤษฎีไวยากรณ์ฟังก์พ เพื่อใช้  
ในระบบการแจงส่วนประโยคด้วยคอมพิวเตอร์ โดยได้เลือกประโยคสำหรับใช้ทดสอบจำนวน 50 ประโยค


การวิเคราะห์เน้นที่การหาโครงสร้างที่แสดงความสัมพันธ์แบบฟังก์พ ทั้งในระดับวากยสัมพันธ์  
คือ การหาโครงสร้างต้นไม้ฟังก์พ และในระดับบรรทัดคำศัพท์ คือ การหาโครงสร้างสายใยความหมาย  
การวิเคราะห์หาโครงสร้างต้นไม้ฟังก์พนั้น อาศัยความรู้ที่สำคัญ คือ เรื่องของหมวดคำและวากยการก  
ของภาษาไทย การกำหนดความสัมพันธ์ในโครงสร้างต้นไม้ฟังก์พนั้น ใช้การเรียงลำดับความสำคัญสาม  
ลักษณะเป็นเกณฑ์ คือ ความสำคัญในแนวลึก ความสำคัญในระยะใกล้ และ ความสำคัญในความเป็นไปได้  
ส่วนการวิเคราะห์หาโครงสร้างสายใยความหมายนั้น อาศัยความรู้ที่สำคัญ คือ เรื่องของการกล่มสัมพันธ์  
และเงื่อนไขการก

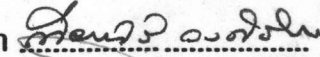
การวิเคราะห์ทั้งในสองระดับนี้ แสดงได้ด้วยกฎการแจงส่วนประโยค ซึ่งเขียนขึ้นเพื่อใช้กับ  
ระบบการแจงส่วนประโยค CUPARSE กฎเหล่านี้จะถูกจัดเป็นชุดกฎต่างๆ โดยที่ในระดับวากยสัมพันธ์  
นั้น ใช้กฎ 10 ชุด ส่วนในระดับบรรทัดคำศัพท์นั้น ใช้กฎ 4 ชุด

ภาควิชา ..... ภาษาศาสตร์

สาขาวิชา ..... ภาษาศาสตร์

ปีการศึกษา ..... ๒๕๓๒

ลายมือชื่อนิติ ..... 

ลายมือชื่ออาจารย์ที่ปรึกษา ..... 

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....



WIROTE AROONMANAKUN : A DEPENDENCY ANALYSIS OF THAI SENTENCES FOR A COMPUTERIZED PARSING SYSTEM. THESIS ADVISOR : ASST. PROF. PEANSIRI VONGVIPANOND, Ph.D. THESIS CO-ADVISOR : ASSOC. PROF. SOMCHAI TAYANYONG. 242 PP. ISBN 974-577-693-9.

This study aims at analyzing Thai sentences by using a dependency grammar. Fifty sentences are selected as corpus.

The analysis yields two types of representation : syntactic dependency tree and conceptual network. Thai categories and Thai syntactic cases are proposed to convert a linear string of sentences into a dependency tree. The construction of syntactic relations on a dependency tree is governed by three sets of priority : bottom-up priority, immediacy priority and probability priority. The dependency tree is then converted into a conceptual network by the use of case mapping and conceptual case constraints.

The analysis is implemented in terms of rules on CUPARSE, a computerized parsing system. Rules for both syntactic and conceptual dependency analysis are written and organized into links. Ten links are used for syntactic analysis and four links are used for conceptual analysis.

ภาควิชา ..... ภาษาศาสตร์  
สาขาวิชา ..... ภาษาศาสตร์  
ปีการศึกษา ..... ๒๕๓๒

ลายมือชื่อนิติ ..... *[Signature]*

ลายมือชื่ออาจารย์ที่ปรึกษา ..... *[Signature]*

ลายมือชื่ออาจารย์ที่ปรึกษาช่วย ..... *[Signature]*



## Acknowledgements

I am indebted to many people who have contributed to this thesis, both directly and indirectly.

First of all, I would like to express my gratitude to my thesis advisor, Dr. Peansiri Vongvipanond, for her valuable suggestions, and her endless patience and thorough supervision, without which this thesis would never have been completed.

I am indebted to all the committee members, Associate Professor Somchai Tayanyong, Dr. Pranee Kullavanijaya, and Dr. Kingkarn Thepkanjana, for their helpful and valuable comments and suggestions. My thanks also go to Dr. Zaharin Yusoff of the University Sains Malaysia for his valuable comments on the future of CUPARSE. I would like also thank to Prima Taveesin for her patient proof reading. My appreciation goes also to all my friends and colleagues in the department for their encouragement. Mr. Jims Wilcox of the English Department helped with my English. I highly appreciate his kind assistance.

Lastly, I would like to express my affection and appreciation for Mr. Suwat Autsawachaichan, who joined me in the design and development of CUPARSE to be used as a testing tool in this analysis. And to my parents and everyone in my family, I offer my heartfelt thanks.



## TABLE OF CONTENTS

	page
Abstract (Thai) .....	iv
Abstract (English) .....	v
Acknowledgements.....	vi
Table of contents.....	vii
Symbolic terms.....	x
Abbreviation terms.....	xi
1. Introduction.....	1
1.1 Rationale.....	1
1.2 Objectives.....	2
1.3 Scope.....	2
1.4 Review of literature.....	2
1.4.1 Parsing.....	2
1.4.2 Semantic network.....	5
1.4.3 Dependency grammar.....	9
1.4.4 Case relation.....	12
2. A dependency grammar for Thai.....	14
2.1 Model of dependency grammar.....	14
2.2 Syntactic component.....	17
2.2.1 Dependency tree (D-tree).....	17
2.2.2 Categorization.....	18
2.2.2.1 Categorization for Thai.....	19
2.2.2.2 Category inventory for Thai.....	21
2.2.3 Syntactic cases.....	24
2.2.3.1 Criteria for syntactic case analysis.....	24
2.2.3.1 Syntactic case analysis.....	28
2.2.3.2 Syntactic case inventory for Thai.....	34
2.2.4 Priority of syntactic relation.....	37
2.2.4.1 Botto-up priority.....	38

	page
2.2.4.2 Immediacy priority.....	39
2.2.4.3 Probability priority.....	40
2.3 Conceptual component.....	44
2.3.1 Conceptual cases.....	46
2.3.2 Conceptual attributes.....	49
2.3.3 Case mapping.....	51
2.3.4 Conceptual case constraints.....	53
2.3.5 Conceptual hierarchy.....	53
2.4 Dictionary.....	55
3. CUPARSE System.....	58
3.1 Introduction to CUPARSE.....	58
3.2 LDCC Module.....	66
3.3 LD Module.....	67
3.4 PPG Module.....	71
3.5 DTC Module.....	71
3.6 CCA Module.....	76
4. Analysis of corpus.....	79
4.1 LD analysis.....	79
4.2 DTC analysis.....	80
4.2.1 D-tree analysis of corpus.....	80
4.2.2 Postulation of rules.....	81
4.2.2.1 Rules for syntactic cases.....	81
4.2.2.2 Priorities of rules.....	83
4.2.3 Links and phases.....	93
4.2.3.1 Grouping rules into links.....	93
4.2.3.2 Rule phases.....	96
4.2.4 Link order.....	98
4.3 CCA analysis.....	99
4.3.1 Components of CCA analysis.....	99
4.3.1.1 Conceptual hierarchy.....	99
4.3.1.2 Conceptual case constraints.....	101



	page
4.3.1.3 Case mapping.....	102
4.3.2 Analysis phases.....	104
4.3.2.1 Subject-object resupplying.....	105
4.3.2.2 Case assignment.....	106
4.3.2.3 Case selection.....	109
4.4 Output of analysis.....	111
4.4.1 Lexical ambiguity.....	113
4.4.2 Category ambiguity.....	114
4.4.3 Nominal prefix.....	115
4.4.4 Case mapping.....	116
4.4.4.1 One to many mapping.....	116
4.4.4.2 Many to one mapping.....	117
5. Conclusion.....	119
5.1 Summary of analysis.....	119
5.2 Suggestions.....	120
5.2.1 Conceptual cases in term of perspectives.....	120
5.2.2 Priorities of relations.....	122
5.2.3 Further development of CUPARSE.....	123
Bibliography.....	124
Appendix A: Sentences corpus.....	130
Appendix B: D-trees and C-nets of corpus.....	132
Appendix C: Dictionary.....	185
Appendix D: LD rules, links, and linkorder.....	192
DTC rules, links, and linkorder.....	193
CCA rules, links, and linkorder.....	208
Appendix E: Grouping of abbreviation terms.....	228
Appendix F: List of Thai words.....	235
Vita.....	242



## Symbolic Terms

$!X \rightarrow Y!$	A relation of head X and depender Y. (no left or right significance).
$!X-R \rightarrow Y!$	R is a case relation of head X and depender Y. (no left or right significance)
$X \rightarrow Y$	A relation of head X and depender Y.
$X < Y$	A relation of head Y and depender X.
$X-R \rightarrow Y$	R is a case relation of head X and right depender Y.
$X < -R-Y$	R is a case relation of head Y and left depender X.
$[X < - \rightarrow Y]$	A relation of head X and right depender Y iff $X \rightarrow Y$ exists. A relation of head Y and left depender X iff $X < - Y$ exists.
$X \rightarrow Y > A \rightarrow B$	Relation $X \rightarrow Y$ has higher priority than $A \rightarrow B$ .
$X \rightarrow Y < A \rightarrow B$	Relation $X \rightarrow Y$ has lower priority than $A \rightarrow B$ .
$R(X, Y)$	R is a case relation of head X and depender Y.
$X_{\Delta}$	The subtree consisting of X as the top node.
$..X...Y...Z$	A lexeme sequence consists of lexemes X, Y and Z with possible intervening lexemes.
#	Sentence boundary

## Abbreviation Terms



ABOUT	About
ABOUTPR	About prep relation
ABST	"ABSTRACT"
ACT	"ACTION"
ADJR	Adjective relation
AFF	Affected
AGT	Agent
ALC	All Left Chain
ANIM	"ANIMATE"
ARC	All Right Chain
ARTIF	"ARTIFACT"
AUX	Auxiliary
Approx	Approximate
BENPR	Benefactive prep relation
BPT	Bottom up priority
BUILD	"BUILDING"
C-Net	Conceptual network
CCA	Conceptual case assignment
CFRM	Case frame
CLSS	Classified
CLSS	Classifier noun
CLSSG	Classifier group
CLSSR	Classifier relation
CMNN	Common noun
CMPL	Complement
CMPLN	Complete noun
CMPNR	Compound noun relation

COMM	"COMMUNICATION"
COMT	Comitative
COMP	Complementizer
COMP	Comparison
COMP	"COMPANY"
CONC	"CONCRETE"
CONCC	Conceptual case
CONN	Connector
CONSTR	"CONSTRUCTION"
CP	Concept
CRDN	Cardinal noun
CSAFF	Constraints on AFF
CSAGT	Constraints on AGT
CSCMPL	Constraints on CMPL
CSINS	Constraints on INS
CSMNS	Constraints on MNS
CSOBJ	Constraints on OBJ
CSPRPT	Constraints on PRPT
Come	Come_Direction
D-tree	Dependency tree
DEFN	Definite noun
DESC	"DESCRIBE"
DET	Determiner
DTC	Dependency tree construction
Demons	Demonstrative
Down	Down_Direction
Dyn_prog	Dynamic_progressive
FOBR	First object relation
FURN	"FURNITURE"
Go	Go_Direction
HM_INTLL	"HUMAN_INTELLIGENCE"

HUMN	"HUMAN"
INANM	"INANIMATE"
INH	Inherited
INS	Instrument
INSTI	"INSTITUTE"
IPT	Immediacy priority
Keep	Keep_state
LAAMD	Left additional modal
LAASP	Left aspect
LAATT	Left attribute
LABEL	Label
LAIRL	Left irrealis
LAMD	Left modal
LAMDR	Left additional modal relation
LANG	"LANGUAGE"
LASPR	Left aspect relation
LATPR	Locative_at prep relation
LATT	Left attribute
LATTR	Left Attribute relation
LD	Lexical disambiguation
LDCC	Load dictionary and create chart
LDET	Left determiner
LDETR	Left determiner relation
LFRTNPR	Locative_front prep relation
LINPR	Locative_in prep relation
LMBD	Left mark boundary
LMDR	Left modal relation
LOC	Location
LONPR	Locative_on prep relation
LRC	Left-Right Chain
LTNSR	Left tense relation

LUNPR	Locative_under prep relation
MACH	"MACHINE"
MAJCAT	Major category
MAN	Manner
MANU	"MANUFACTURE"
MAN_STS	"MANNER_STATUS"
MEAS	"MEASURE"
MFOBR	Case mapping of FOBR
MINCAT	Minor category
MNS	Means
MODAL	Modal
MORPH	Morphological form
MPOSSPR	Case mapping of POSSPR
MSUBR	Case mapping of SUBR
MTOPR	Case mapping of TOPR
MWITHPR	Mean_with prep relation
N	Noun
NOM	Nominalized
NUMB	Number
NUMB	"NUMBER"
NUMR	Number relation
Non_spec	Non_Specific
OBJ	Object
ORGA	"ORGANIZATION"
Oblige	Obligation
Opport	Opportunity
PARTN	Partner
PATT	Pattern
POSS	Possesor
POSSPR	Possesive prep relation
PPG	Possible paths generation

PPT	Probability priority
PREF	Prefix
PREP	Preposition
PREPN	Preposition (for noun)
PREPV	Preposition (for verb)
PROC	"PROCESS"
PROD	"PRODUCT"
PRON	Pronoun
PRPN	Proper noun
PRPT	"PROPERTY"
PRPT	Property
PRPT_STS	"PROPERTY_STATUS"
PRTOF	Part_of
PS-tree	Phrase structure tree
PT_BUILD	"PART_BUILDING"
PT_MACH	"PART_MACHINE"
QUAT	Quantity
RAASP	Right aspect
RAATT	Right attribute
RASPR	Right aspect relation
RATT	Right attribute
RATTR	Right attribute relation
RDET	Right determiner
RDETR	Right determiner relation
REL	Relator
RELMS	Relative missing
RFROMPR	Range_from prep relation
RLC	Right-Left Chain
RMBD	Right mark boundary
RTOPR	Range_to prep relation
SOBR	Second object relation

SOR	Source
SUBR	Subject relation
SYNTC	Syntactic case
Stat_prog	Static_progressive
TAR	Target
TIM	Time
TIM_B	Time_Begin
TIM_E	Time_End
TOPR	Topic relation
UPCP	Upper concept
Up	Up_Direction
V	Verb
VADJ	Adjective verb
VCMN	Common verb
VEQU	Equative verb
VEXI	Existential verb