

Chapter 1

Introduction

1.1 Introduction

This thesis is broadly concerned with the nature of data in linguistics and more specifically with the role of grammaticality judgments offered as evidence in recent Thai linguistic research. Linguistic research of any type involves data but the definition and meaning of what is considered “data” varies depending on the approach taken to the study of language within the broader discipline of linguistics. Data used as evidence for hypothesized syntactic or semantic structures often consists of simply the self-generated sentences offered by the researcher doing the research. Preliminary steps to verify the grammaticality or acceptability of these sentences are, as a matter of course, not considered necessary. However in recent years research has been done that shows the inadequacy and unreliability of single informant grammaticality judgment data (Featherston, 2007; Schutze, 1996; Cowart, 1997).

Research on Thai serial verbs has yielded a range of acceptability judgments that draw into question the problematic nature of clear grammaticality judgments. As this phenomena became apparent in the researcher’s own investigation, it became desirable to check the data underlying several major works on Thai serial verbs and see how consistently the grammaticality judgments of a larger pool of subjects correlated with the grammaticality judgments that those works were based on. The Thai language along with many other languages in this region like Chinese, Khmer, Vietnamese and Lao exemplifies “an extreme of pragmatically-oriented grammar” (Enfield, 2007: 272). One important consequence of this fact in regard to testing sentences for grammaticality is that a decontextualized sentence can be open to a wide range of interpretations or to no immediate accessible interpretation without appropriate supplemental contextual information. A second fact about languages in this region that complicates syntactic analysis in general is that sentences are frequently one clause made up of multiple verbs. As will be discussed in detail in Chapter 5, this structural feature has resisted efforts of linguists to come to analytical agreement or even descriptive agreement in characterizing the

phenomena of serial verb constructions. The resistance to description or analysis may stem from the fact that sentences in Thai simply do not have to have a main verb as a clausal core feature. The syntactical role of the main verb is to mark clause boundaries but serial verbs in Thai can cohere as one event, describe a temporal series of events, an event and an evaluation of that event or an event and result of that event.

This investigation was undertaken using a basic written survey with bilingual native Thai speakers and personal interviews with monolingual Thai speakers. The results show that the purported acceptability or non-acceptability of some of the presented tokens failed to be recovered when tested with native speakers. In some cases sentences when tested did not fall into the predicted categories of acceptable or unacceptable for a majority of my informants. Additionally, claimed semantic contrasts between certain paired sentences were not found.

1.2 Thesis overview

In Chapter 2, I discuss three broad perspectives to the study of language that have been termed Externalist, Emergentist, and Essentialist approaches to linguistics. Linguistics has emerged historically as a field of study with broadly different conceptions as to what the subject of study should be, what the goal of study should be and what counts as explanation. The concepts of grammar in each of the three approaches are contrasted and the special status of grammaticality judgments in the Essentialist approach emerges as the main problematic issue regarding the status of empirical data used as evidence in theory construction.

In Chapter 3, I present the results of a survey that I performed using questionnaire methods and interview techniques to check the acceptability claims of 20 sample Thai sentences. Each of these sentences has appeared in the literature as tokens that supported specific theoretical claims in Thai linguistic research. The claims are reviewed briefly before the results of the survey are given for each claim. The sentences were surveyed for acceptability by native speakers, meaning recovery of the English free translations and in some cases claimed meaning contrasts in paired sentences. The results of the survey revealed that concerns about the empirical basis of much linguistic research are warranted.

The problem of reliability of grammaticality judgment data and the purported difference between acceptability and grammaticality as distinct concepts is discussed

in some depth in Chapter 4. I tested English sentences with 24 native English speakers that had been offered as evidence for syntactical claims to show the reader the difficulty of obtaining reliable consistent data which can be considered empirical. The normative nature of grammaticality judgments as well as the normative or prescriptive notion of grammar itself is an issue at the core of linguistic research and I review the Essentialist perspective on this issue. Recent empirical research on English syntax by computer linguists developing Natural language processing Technology reveals that the results of Essentialist syntactical research is of little use in this effort.

In Chapter 5 specific issues concerning linguistic theory and the languages of Southeast Asia are discussed. The languages of SE Asia seem to be defined in terms of what their grammars lack rather than how their grammar actually structures their languages. The lack of a core verb is perhaps the issue that seems most perplexing to the Eurocentric view of sentence structure. In this chapter the issues of serial verbs and the confusion surrounding efforts to fit these structures into an analytical framework is reviewed. I return to the data in Chapter 3 and show in detail how certain sentences were problematic due to semantic confusion. I offer the notion of a semantic versus a syntactic grammar and show how sentences in Thai lack the syntactic grid of an English sentence where grammatical morphemes leave slots for open class word substitutions.

1.3 Introduction to the problem

The etymological meaning of the word “data” is “(things that are) given”. When one researcher takes the analysis of another researcher as an example or evidence for argumentation, analysis has now become “data”. Sentences can be used as data and an analysis of a sentence can be data as well. Something is not, in and of itself data, since data is connected to a function that it serves (Lehmann, 2004: 4). Data can be more or less abstract as the two cases above illustrate. In this thesis evidence presented as data in the research of others was checked for the most basic empirical claims and not the abstract claims made for the sentences as evidence for syntactic and semantic argumentation. To make this point more clear, any sentence presented by a linguist as evidence has in Lehmann’s terms become “semiotic object of a higher order”. It is an expression in language coupled with a statement of the meta-language which predicates a certain property over that object. The question posed and answered was, can these sentences stand up to a check with native speakers for

acceptability, meaning translation and contrast claims between certain pairs of sentences?

1.4 Basic introduction to Thai

Thai is the most studied and described language in the family of Tai-Kadai languages. In fact, over half of the approximately 90 million speakers of Tai-Kadai languages speak Thai as a first or second language (Diller, 2008: 31).

In Thailand the linguistic situation is extremely complex and by far the best explication of the linguistic situation in Thailand is Smalley's *Linguistic diversity and national unity: Language ecology in Thailand* (1994). Thais themselves are often unaware of the fact that a majority of Thais learn Standard Thai in school. Thais with a high school education will say that they are Thai and that they speak Thai. Further questioning is usually required to elicit the fact that they also speak a regional language and that language is actually their first language. The prestige of Standard Thai is unquestioned by all Thais. In fact as Smalley points out, Northern Thais will insist that their language, Khammuang, is not a language in the sense that Thai is a language. It is rather, "village speech", or just a version of Thai that country people speak. An education, a career, and economic success are dependent upon mastery of Standard Thai. The language variety of Thai that is studied in this thesis is Standard Thai but differentiating Standard Thai from Central Thai is difficult. An interesting definition of what Standard Thai is perceived to be is quoted from Beebe (1974) in Smalley (1994).

Standard Thai has been called a national language, a regional dialect, a language for use in schools, an equivalent of Central Thai, an equivalent of Bangkok Thai, an equivalent of Central or Bangkok Thai, a language not equivalent to, but closely resembling Central Thai, a prestige dialect of educated speakers regardless of origin, and a model of what Central Thai is supposed to be (Beebe, 1974: 74).

The Central dialect, to which Standard Thai belongs, is one of the four major dialects and has 20 to 25 million speakers. The other major dialects are the Northeastern dialect (Isaan or Lao) with about 23 million speakers, the Northern dialect (Kham Muang, Lan Na or Yuan) with 6 million speakers, and the Southern dialect with 5 million speakers (Summer Institute of Linguistics website).

Besides these major dialects, a number of related languages are spoken in the country. Almost all of these languages belong to the Southwestern branch of Tai (Li, 1977). Speakers of these Tai languages comprise about 90% of the whole population. Non-Tai languages include Khmer (Austro-Asiatic), Malay (Austronesian), Karen (Tibeto-Burman), and Hmong (Hmong-Mien). The 2000 census reports that 2.3% of the population speaks Khmer and 2.3 % Malay. In addition, in cities and towns southern varieties of Chinese such as Teochiu (or Swatow) and Hakka are commonly found. Languages of the Southwestern branch of Tai are found not only in Thailand but also in Laos, northern Vietnam, Myanmar, India, and also in southern China. The other two branches are the Central and Northern branches. The Central branch includes languages spoken in northern Vietnam and southern China (e.g. Nung, Tay). The Northern branch includes other languages of southern China (e.g. N. Zhuang, Bouyei, Seak). The Tai language family with these three sub-groups is related to other sister and parent branches to make up a larger language stock called Kadai or Tai-Kadai.

The figure below shows the position of Thai in a language family tree. This tree is from *The Tai – Kadai Languages* (Diller, Edmundson and Luo, 2008: 7) where it is described as “a tentative diagram for reference”.

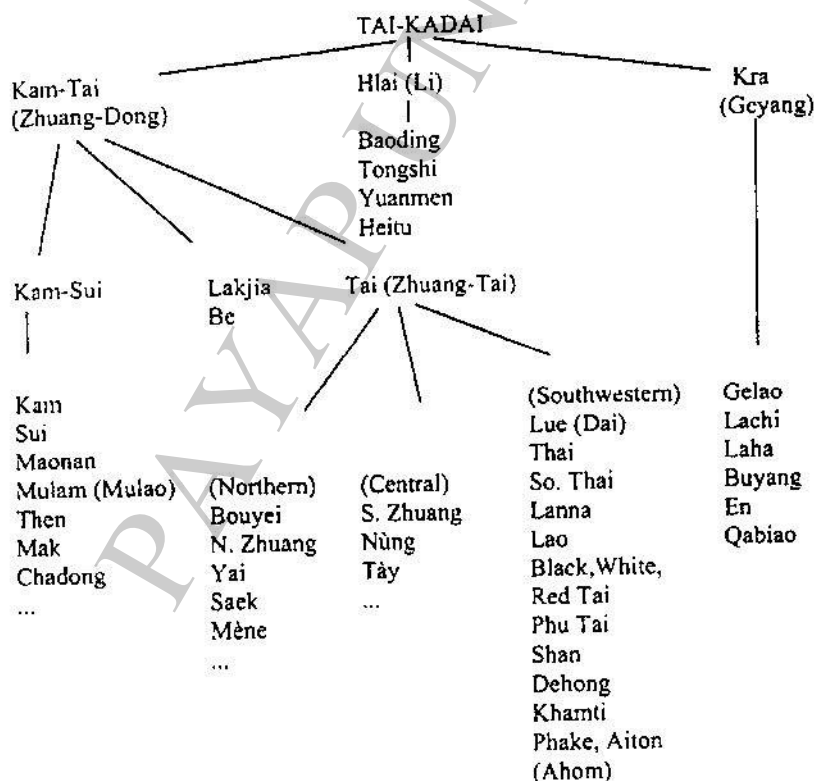


Figure 1: Tai Language Family

1.5 Summary of Methodology

The core data for this thesis comes from taking examples of serial verb sentences from the linguistic literature on Thai and asking native Central Thai speakers if they find these sentences acceptable. Both questionnaire methods and interviews were used to gather the data for this thesis. Respondents were asked to rate the sentences for acceptability as OK, confusing but OK, unacceptable, or uninterpretable. I asked these bilingual students to also translate the sentences in English and make comments on the sentences regarding the structure or vocabulary of each sentence. In many cases where the sentences were judged to be unacceptable, reasons for their unacceptability were discussed in subsequent interviews. Approximately half of the informants were fluent to some degree in English and many of their comments appear in Chapter 3. Others were Thai speakers and tapes were made and notes taken on their responses during the interviews.

1.6 Contribution of the thesis

The use of linguist self-generated sentences used as data has been controversial for decades as will be discussed in detail in Chapters 2 and 4. The criticism has focused on Essentialist (generative) methodology in particular but the practice is still being defended by linguists. Jackendoff believes that “it would cripple linguistic investigation to require that all judgments of ambiguity and grammaticality be subject to statistically rigorous experiments on naive subjects” (Jackendoff, 2010). But in this modern age where information is collated, processed, sifted, and stored at a lower cost than ever before, can this lack of interest in the details and specifics of data and data collection be upheld in a field which purports to be an empirical science?

Critics of linguists’ use of intuitive data are not recommending that detailed statistical analysis be done on informant grammaticality judgments. Featherston argues that at an utter minimum for the basis of advancing a linguistic argument multiple informants along with multiple lexical variants of the structures investigated should be required and that this information should be presented as part of the research (Featherston, 2007: 10).

In Chapter 3 I present in detail the actual reality of attempting to access the postulated linguistic competence of Thai native speakers asked to judge acceptability of 20 sentences. The result reveals in detail that often informant judgments are not

reflexive, binary or unproblematic. The more experience one has with this process, the less one can believe that syntactic and semantic levels of interpretation are cognitively separable.

1.7 Limitations and scope

The concept of data in the various approaches to linguistics is a wide domain to explore and much more could be brought out especially with regard to the revolution in data processing capability that is occurring today in corpus and computer linguistics. Trends indicate that our recent ability using computers to examine, sort, collate and analyze language will revolutionize the subject of natural language research. My research was aimed only at answering a limited number of basic questions regarding the reproducibility of basic data. What this thesis does not do is:

1. Attempt a re-analysis of the original work.
2. Argue the correctness of the original syntactical or semantic claims.

1.8 Similar research

The Thai word *wāa* originally meaning “say” can serve as a complementizer for verbs of expression, perception and judgment as shown in the example below.

(1)

Nuan khít wāa (zero) thǎy paj roŋ rian wan ní
Nuan think that - should go school day this
“Nuan thinks that (he) should go to school today”

In answering the question, “Can the zero element in this sentence refer to a subject outside of the sentence or is this ambiguity not allowed?” different answers are given by two different Western trained Thai linguists in their PhD dissertations. Pingkarawat (1989) claims that this is a case of syntactic binding and sentences like #0 are not ambiguous while Hoonchamlong (1991) says that referents outside the sentence are allowed.

This intriguing situation is discussed in detail by Diller and Khanittanan (2002) who attempt to test this disagreement by surveying other Thai native speakers. The point is made that this is not a trivial disagreement since questions of zero reference in languages such as Thai, Lao, and Chinese are of particular interest since generative approaches have sought to formulate a universal taxonomy of zero elements (e.g. pro, trace, PRO, and the like) and to establish cross-linguistic parameters to account for the distribution of zero elements. Diller and Khanittanan attempted to test sentences with zero elements with 45 native speakers in an effort to derive a possible taxonomy of zero elements for Thai. Their conclusion was that “success in imagination” was responsible for many of the positive acceptability judgments and they raise the critical question of whether creating a taxonomy of zero elements using grammaticality judgments is, in fact, a possible goal.