

SYNTACTIC DUALITY IN CLASSICAL TAMIL POEMS

JEAN-LUC CHEVILLARD

“Since the punctuation of compounds is my own responsibility, it is no doubt up to me to try to justify the practice. [...] So far as the grammatical analysis itself is concerned, I am anxious not to give the impression of imagining that experienced Sanskritists stand in the same need of it as an aid to interpretation as I do myself. One might nevertheless urge its merits as a simple convenience to all readers, like word-division, the use of which does not imply that scholars would be unable to separate the words out for themselves. However my case rests on quite different grounds, those supplied by Hertel to justify word-division itself, namely that its use provides the simplest form of running commentary. To justify his choice of readings, an editor must make clear how he construes his own text.”

(M. Coulson, *Mālatīmādhava*, introduction, p. liii)

When one attempts to make a linguistic description of a dead language, some limitations have to be kept in mind:

- **Limitation 1** : one works on a finite set of data (a closed corpus of texts), and that text has to be handled by the methods of philology¹ (and it can have variants).
- **Limitation 2** : one cannot check that data with native speakers and obtain feedback, and great parts of the data may be poorly understood (or not at all).

If that language has possessed at one point in time an indigenous grammatical, lexicographical and/or commentatorial tradition, the second limitation may seem to be partly overcome, because a one-way flow of information can be established with a, hopefully, native source of linguistic knowledge, provided one is aware of another limitation:

- **Limitation 3** : the perspective of a traditional grammarian, or lexicographer, or commentator, does not necessarily coincide with that of a modern descriptive linguist.

If, additionally, the texts one studies are literary/poetical texts, there appears a supplementary difficulty which could be

¹ For sharing with me her insights concerning many points discussed here and for reading several successive versions of this paper, I would like to express here my special thanks to my philologist colleague, and learned friend, Eva Wilden. All errors are of course mine.

formulated thus:

Limitation 4: Can a language used in literary/poetical texts be said to have had native speakers? and should it be recognized as a natural language or as an artificial creation?²

Defining perfect guidelines for avoiding all the pitfalls which are the consequences of these four limitations is beyond the scope of this paper, which solely concerns itself with one case study concerning such a language, namely Classical Tamil. However I shall try to keep them in mind.

Classical Tamil: the language of a corpus

Many present-day speakers of Tamil would probably be shocked to see Classical Tamil described as a "dead language": one often meets with statements that Tamil is at the same time a living and a classical language. Their perception is shaped by the fact that a sizeable part of the vocabulary of Classical Tamil is still in use in the H (or formal) variety of Modern Tamil, and also probably by the fact that memorizing of some CTamil texts (like the *Kural*), at least partly, is traditionally done in schools. Moreover, a number of bridges between the past and the present exist, some of them being commentaries, glosses, translations from CTamil to more recent varieties, and some of them being scholars who have been specially trained in reading the ancient texts and their ancient commentaries, having received their knowledge from earlier scholars. Nevertheless, it is improbable that there has been a perfect continuous line of transmission of the understanding of classical texts, and even traditional scholars cannot play the role of quasi-native speakers for Classical Tamil (henceforth CTamil), although some of them can certainly compose poetry in neo-classical Tamil.

The CTamil texts which we shall examine here, are poems which were composed at dates difficult to determine. The main vehicles through which these poems have been handed down to us are anthologies, compiled at dates which are also not well known.³ One of the problems that face us is to determine whether some anthologies are, as a whole, later than others. Another problem, which has possibly been examined less often, is to examine whether inside each anthology there is linguistic homogeneity. An improvement of our knowledge in this respect might be

² That language could possibly be described as a creative extension of a natural language, obeying a set of constraints, that differs [in a dual/parallel way] from those for another language with which it is in symbiotic relation in a diglossic situation. For the Tamil diglossia, see Britto [1986]. For other perspectives on diglossia, see the volume edited by Bh. Krishnamurti [1986] *South Asian Languages: structure, convergence and diglossia*.

³ For a discussion of the chronology, see Wilden [2002] [WZKS, XLVI (105-133)]

a way of improving also our grasp on their internal chronology. With these objectives in mind, a program of systematic observations of the linguistic characteristics of several CTamil anthologies is currently being conducted⁴ and concentrates for the time being on a set of 3 early anthologies of love poetry: the *Kuruntokai* (KT, 401 short poems, 4 to 8 lines), *Narriṇai* (NA, 400 poems, 9 to 12 lines) and *Akanānūru* (AN, 400 long poems, 13 to 31 lines). The results presented here will mainly be drawn from 4 corpora, listed here by ascending number of lines.

- CK [152 lines] (a subset of KT): the 38 poems having 4-lines inside the KT
- NE [658 lines] (a subset of AN) consisting of the 40 *neytal* poems⁵
- KT [2504 lines] (the complete anthology)
- NA [4180 lines] (the complete anthology)

A dual Syntax: M-syntax vs. U-syntax

The main point which I would like to emphasize in this study, is that there is a duality in the syntax of these texts contrasting two poles, which I shall characterize as a **marked syntax** (or M-syntax) opposed to an **unmarked syntax** (or U-syntax), and that most of the poems that belong to the corpus bear traces of these two syntaxes, in ways which I shall try here to quantify. The roots of that phenomenon probably lie in the diglossic situations that have been existing at the various periods when the poems were composed: M-syntax is probably rooted in the “vernacular speech” that was in use at that time, although the poetical constraints certainly restricted/filtered the spectrum of what was put inside the mould; on the other hand, U-syntax would draw its strength/impetus/efficiency from the poetical craft of learned men vying with (or emulating) each other in the use (and the enrichment) of a stock of formulae, where the use of what traditional CTamil grammar calls *tokai* (or *tokuti*), usually translated as “compounds”, is frequent.

The terminology which I shall use in describing the CTamil texts is ultimately based on the one found in the oldest preserved CTamil grammar, the *Tolkāppiyam* (T), subsequently reworked by later grammars like the *Nannūl* (N), and partly adapted into western languages

⁴ This effort, by my colleague E. Wilden and myself, is itself part of a wider Pondicherry “Sangam Project”, that has the ambition to cover the whole of CTamil literature.

⁵ The *neytal* poems are devoted to the sea-shore setting. A critical edition and translation project is currently underway (Wilden & Chevillard, forthcoming). A full edition and translation of the 400 poems of AN [totalling 7151 lines] is also planned.

(Latin, English, etc.) by missionaries (most notably Beschi) who had acquainted themselves with traditional grammars. As seen through that terminology, there are in CTamil, to begin with, two main inflected parts of speech, the noun (*peyar*) and the verb (*vinai*, *tolil*),⁶ but that basic terminological dichotomy allows for the presence of other categories. For instance, the forms that we could consider as predicative adjectives, are treated by the T as occurrences of special verbs, called *kurippu vinai* "suggestive verbs", morphologically defective with respect to the expression of tense.

My description of M-syntax will be quite incomplete. The reason is that it is much more complex than U-syntax. I shall in fact concentrate on those features in M-syntax which have a dual feature in U-syntax. One such area, in the case of the verbal paradigm, is the use of the two main types of *peyar-eccam* "relative participles", designated here as V^{PE1} and V^{PE2} , which is echoed in U-syntax by constructions using plain verbal roots, designated here as V^0 , which could be considered as an unmarked relative participle. Both will be presented, in a preliminary fashion, in the next paragraph.

Argument positions around a relative participle

The *peyar-eccam* forms are so called because they are (syntactically) "incomplete" (*eccam*) forms, which have to be completed by a *peyar* "noun", which immediately follows it. We shall refer to this noun, N_1 , as the "head noun" and shall consider it as the most prominent syntactic argument for the participle. Many relative participles have other arguments, N_2 , N_3 , etc., standing on the left, and a typical template for these constructions can be represented as:

$$\{((... N_3) ... N_2)V^{PE}N_1\} \text{ (where } V^{PE} \text{ could be } V^{PE1} \text{ or } V^{PE2})$$

The difference between the V^{PE1} and V^{PE2} types is temporal (or aspectual) but syntactically they behave in the same way, although we shall see later that the most frequent semantic roles for their arguments are not the same. I shall give a first series of examples with the verbs *utirtal* "to fall" and *aṭutal* "to kill". Their PE1 forms are *utirum* "which-falls" and *aṭūum*⁷ "which kills". Their PE2 forms are *utirnta* "which-fell" and *aṭṭa* "which-killed". The English translations given here are approximations, because the head noun can have various roles with

⁶ In their main occurrences in non-technical literature, *vinai* is "action" and *tolil* "activity".

⁷ There is another variant : *aṭum*.

respect to the PE, as shall be seen in some of the following examples, where the head noun is underlined

- *utirnta kaṇi* (fall^{PE2} ripe-fruit) "fruit which fell" [AN 380-4/5]
- *utirum ... malar* (fall^{PE1} flower) "flowers which fall" [AN 199-2/3]⁸
- *aṭṭa mallar* (kill^{PE2} warriors) "warriors which killed" [KT 34-5]

I shall refer to such examples as instances of $\{V^{PE1}N\}$ or $\{V^{PE2}N\}$ constructions. These examples, which are "real" examples extracted from the CTamil corpus, are simply the most elementary ones, inside a class of examples where we most often find more than one nominal argument, as in the following two examples, which are instances of, respectively, $\{N_2V^{PE2}N_1\}$ and $\{N_3...N_2V^{PE1}N_1\}$ constructions.⁹

- *kāy_{N2} utirnta [paṇi kulai]_{N1}* (unripe-fruit fall^{PE2} green bunch) "green bunch from which unripe fruits have fallen" [AN 21-20]¹⁰
- *[puli ērrai]_{N3} ... kaṭiru_{N2} aṭūum ... verpu_{N1}* (tiger male elephant-bull kill^{PE1} mountain) "the mountain where the male tiger kills the elephant-bull" [NA 36-1/4]

It should also be noted here that, although the inflectional paradigm of the noun is primarily described in terms of case by traditional CTamil grammarians, the multiple nominal arguments N_1 , N_2 , N_3 that appear in these examples are not morphologically marked and this is quite typical of the language of that corpus, where case morphemes are not at all frequent.¹¹

This information being given, I shall now give parallel examples for the *vinait tokai* "verbal compound", which results, as already noted, from the combination of a verbal root V^0 and of a noun. In the case of the verbs *utirtal* and *aṭatal*, the V^0 are *utir* and *aṭu* and we can see them used in the following examples, combined with a noun,

⁸ In order to keep this initial presentation simple, the other specifiers of *malar* "flower" are not reproduced here.

⁹ I have used square brackets to indicate those nominal arguments that are themselves complex, like *[paṇi kulai]* and *[puli ērrai]*. It has also to be noted that, even though the examples are more complex than the preceding ones, they are simplified, as can be seen from the use of "...".

¹⁰ The head noun has for its referent the starting point of the action of falling.

¹¹ In the real picture of noun morphology in CTamil that emerges from a corpus-based study, the most important distinction is the one between the unmarked and the oblique forms.

- *aṭu pōr* (kill^{V0} war) “deadly battle” (NA 39-9)
- *aṭu piṇaṇ* (kill^{V0} corpse) “killed corpse” (NA 329-2)
- *utir vī* (fall^{V0} flower) “flowers that drop” (AN 99-7)

Just like the V^{PE1} and V^{PE2}, the V⁰ (or plain verbal root) is often found with two arguments, like what we have in the following example, which I shall describe as an instance of {N₂V⁰N₁}:

- *ikal aṭu munpiṇ* (enmity^N kill^{V0} strength^{N-obl.}) “[tusks having the] strength to kill enemies” (AN 172-9)

Analysis of a two-lines sample text

In order to explain more precisely what I refer to when talking about a dual syntax, I shall start with the analysis of a 2-lines passage extracted from a 15-lines poem in the AN anthology. This will also serve as an illustration of the difficulties inherent in dealing with classical poetry, and this is why I shall first give it in its metrical form, accompanied by an unpublished English gloss going back to the late V. M. Subramanya Ayyar (henceforth VMS).

*mutumarat turaiyu muravuvāy mutupuṭ
katumenak kulaṟuṇ kaḷutuvalaṇ karainā!* [AN 260, 12-13]

“At midnight when the demons start to roam about, the old rock-horned owl which has a broken beak and is always residing in the old tree, ..., suddenly cries.” (VMS)

If we separate the metrical feet into words and additionally dissolve the sandhi, while keeping as an information the fact that there was originally a sandhi change¹², we obtain a text, which is the following, in which I have isolated 14 strings.

*mutu marattu{S^{aut}} uraiyum{S^{aut}} muravu vāy mutu puḷ{S^{?sign}}
katum-ena{S^{aut}} kulaṟum{S^{aut}} kaḷutu valaṅku{S^{aut}} arai-nā!*

¹² This information must not be lost, because it can have consequences on how one must interpret the syntactical relations. However, many sandhi rules can be described as “automatic”, and they shall be indicated here by {S^{aut}} but these “automatic” sandhi rules are most of the time not applied by modern editors, which might give the impression to the reader that they are optional. The significant sandhis shall be indicated by {S^{sign}} or by {S^{?sign}} (when there is a doubt).

Using as exponents grammatical tags, intended as short-cuts¹³ into the morphology of Tamil and soon to be explained, and giving the meanings, the 14 strings reduce to 12 items, because *katum-ēṇa* can be analysed as an ideophonic expression, containing the particle *ēṇa* as a constituent, (see Chevillard 2004: 416) and because *arai-nāl* “midnight” is best treated as an unanalysable idiom, although *arai* “half” and *nāl* “day” could exist as stand-alone nouns. The 12 items are as follows:

old^{A0} tree^{N-OBL} reside^{V-PE2} broken^D beak^N old^{A0} bird^N
 suddenly^{ideoph-ēṇa} cry^{V-PE2} demon^N roam^{V0} midnight^N

For reasons which are ultimately rooted in the traditional grammatical analysis of CTamil “compounds”, three phrases inside the text have been underlined. They constitute what I designate as the domain of U-syntax (unmarked syntax), and I shall call them “U-components”. A further step in their analysis, which shall shortly be explained, can be accomplished by adding some punctuation marks, adapted from Coulson [1989]:¹⁴

- (old^{A0}:TREE) “old tree”
- ((broken^D:beak^N)-(old^{A0}:BIRD^N)) “old bird with a broken beak”
- ((demon^N-roam^{V0}):MIDNIGHT^N) “at midnight, when the demons roam about”

For each of the three U-components, the last item, which has been marked here by small capitals, is designated as “the head” of the U-component. When we replace inside the original text each of the U-components by its head, the three heads being “tree”, “bird” and “midnight”, we obtain a reduced text, which is:

¹³ Among the tags that appear as exponents, those starting by V (resp. A or N) indicate forms belonging to a verbal (resp. adjectival or nominal) paradigm. The items with a “D” tag are isolated items which cannot easily be ascribed to a regular verbal, adjectival or nominal paradigm, but which are in the position of determiners. A more thorough examination of CTamil lexicon might however possibly provide ground for considering them as belonging to the paradigms of defective verbs, nouns or adjectives. This is to say that “D” is also a “waiting-room” grammatical category.

¹⁴ In this system (Coulson [1989] pp.1-11), “A;B” is a co-ordinative compound (*dvandva*), “A:B” is a descriptive compound (*karmadhāraya*) and “A-B” is a dependent compound (*taipuruṣa*), both descriptive and dependant compound being considered as “determinative compound” (the wider acceptance of *taipuruṣa*). Additionally, another type of compound, the *bahuvrīhi* type, can be formed on top of these 3 types and is noted by an underlining: “A_·B”, “A_·B” and “A₋B”. Finally, there is special notation “A_B” for prepositional compounds (*avyayībhāva*), a special type of exocentric compound.

tree^{OBL} reside^{V-PE2} bird^N suddenly^{ideoph-ena} cry^{V-PE2} midnight^N

This reduced text constitutes what I designate as the domain of M-syntax (marked syntax) and it can be translated thus:

“at MIDNIGHT, when the BIRD that resides in the TREE suddenly cries”.

The pivotal elements, in the building of the reduced phrase with “marked syntax”, are, in this case, 2 verbal forms, *uraiyum* “reside^{V-PE1}” and *kularum* “cry^{V-PE1}”, which are the *peyar eccam* (“relative participles”) of the *ceyyum* type of the verbs *uraital* “to reside” and *kularutal* “to howl” (see above).

If we now combine the domains of these two syntaxes, we obtain the full text, where the head nouns are in capitals:

“at MIDNIGHT-when-the-demons-roam-about, when the old-BIRD-with-a-broken-beak that resides in the old-TREE suddenly cries.”

Compounds in CTamil: the building blocks of U-Syntax

Lest my presentation of U-syntax and M-syntax might appear as an arbitrary decision of mine, I shall now try to briefly explain the standard “theory” on the compounds which is found in the native CTamil grammars, starting with T.¹⁵ The basic idea, which was certainly adapted from the Sanskrit grammarians’ concept of *samāsa*, is that one can meet with 6 types of complex items called *tokai*-s¹⁶ (see TC412c) which “have the behaviour of a single [inflected] word (*oru col*)” (TC420c), although they contain several components (see TC419c). However, the author (or the authors) of T, which is the most ancient CTamil grammar, seem(s) to have been hesitant to make precise statements on the nature of the

¹⁵ For a detailed exposition, the interested reader who reads French can consult Chevillard [1996] (pp. 491-513).

¹⁶ The word *tokai* itself could be explained in two ways. As Cēṇāvaraiyar, one of the commentators of T puts it: “Les maîtres sont de deux partis : a. ceux qui disent que [les composés] sont [appelés] *tokai* «composés-[à-élision]» du fait que s’élide (*tokutal*) un morphème casuel (*vēṇṇumai ~urupu*), un morphème de comparaison (*uvama ~urupu*), [la particule coordinative] *um*, une finale verbale (*viṇai+col= īru*) ou une finale de mot de qualité (*paṇṇu+col= īru*); et ceux qui disent que [les composés] sont [appelés] *tokai* «composés-[somme]» du fait que sur une [unique] valeur, deux ou plusieurs mots, étant perçus comme une unité (*orṇumai+ paṇṇutal*) sans qu’il y ait fissure (*piḷavu-paṇṇutal*), sont liés (*iyaital*) entre eux.” (TC412c, as translated in Chevillard [1996]). According to Cēṇāvaraiyar, the correct interpretation is the second one.

components to be found inside a *tokai*,¹⁷ although devoting 10 sūtra-s to a presentation of the 6 types of *tokai* which he/they recognize(s) and of their properties. The traditional list of *tokai* found in sūtra TC412c, is the following:

1. *vērrumait tokai* “case compound” (characterized in TC413c),
2. *uvamat tokai* “comparison compound” (characterized in TC414c),
3. *vinaiyintokai* (later called *vinait tokai*) “verbal compound” (lit. “action compound”) (characterized in TC415c),
4. *paṇṇin tokai* (later called *paṇṇut tokai*) “adjectival compound” (lit. “quality compound”) (characterized in TC416c),
5. *ummait tokai* “copulative compound” (lit. “compound” [with] *ummai* [for its value], where *ummai* designates the coordinating clitic *-um*) (characterized in TC417c),
6. *aṇmolit tokai* “compound [with] a non-word”¹⁸ (characterized in TC418c)

We have to rely on the commentators for illustrations, because none are found in the T itself, which contains only characterizations and enumerations. I shall not however reproduce here all the examples they give, but shall also draw examples, as far as possible, from the CTamil corpus and concentrate my efforts on the three types of *tokai*-s most frequently met with, which are the 4th, 3rd and 1st in this list.

One fairly frequent and simple type of *tokai* is the *paṇṇut tokai* “adjectival compound”.¹⁹ A standard example given by commentators is *karuṇ kutirai* (TC410i) “black horse”. In the sample text which we have examined, this type was represented in *mutu marattu* (old^{A0}:TREE^{N-OBL}) and *mutu pul* (old^{A0}:BIRD^N). The head (or second element) is a noun. The first element can be a plain adjectival root, here tagged A⁰; it can also often be an adjectival form with an *-m* or an *-um* ending.²⁰ I shall punctuate them with a colon, imitating Coulson’s representation of Sanskrit descriptive determinative compounds. The example *karuṇ kutirai* (black^{A^{um}}:horse^N) will thus be seen as an instance of A^{um}:N.

¹⁷ The commentators excuse him by saying that the T does not deal with the topic of *oru-moḷip puṇarcci* “internal sandhi”.

¹⁸ These are usually identified with the Sanskrit *bahuvrīhi*-s. The Sanskritized grammar *Vīracōḷiyam* even calls them, in a direct loan translation, *pala nel camācam*-s (Vi_45).

¹⁹ I have already studied that topic in Chevillard [1994].

²⁰ The class of adjectives is not completely homogeneous. The most convenient way of describing the typical adjectives seems to be to consider them as defective verbs, uninflected for tense.

Another very frequent type of *tokai* is the *vinait tokai* “verbal compound”. A standard example given by commentators is *kol yānai* (kill^{V0}:elephant^N) “murderous elephant”. The head (or second element) is a noun. The first element is a plain verbal root, tagged V⁰. The template is thus {V⁰:N}, where I use the same punctuation as for *paṇput tokai*. In the sample text which we have examined, there is indeed a V⁰, but it is found in combination with 2 nominal arguments, in ((demon^N-roam^{V0}):MIDNIGHT^N) “at midnight, when the demons roam about”. It could thus be said to be an instance of a {(N₂-V⁰):N₁} template. The commentators of T quote one example that seems to fit into this template; however they do not do so while explaining *vinait tokai*, but in connection with a T *sūtra* about ambiguous verbal constructions (TC95c). The example given is *puli kol yānai* [(tiger^N-kill^{V0}):elephant^N], which, according to them, can have two interpretations: “elephant which killed a tiger” and “elephant which a tiger killed”. A corpus based examination reveals that V⁰ forms are in fact used more frequently in {(N₂-V⁰):N₁} constructions than in {V⁰:N}. We shall come back to these constructions later. I shall refer to them as “extended verbal compounds”.

The third type of *tokai* which we shall present here is the *vērrumait tokai* “casal compound”, in which a head noun N₁ combines with another noun N₂ in a {N₂-N₁} template. Commentators give series of 6 illustrations for this type of *tokai*, and explain that there is a casual relationship between the two nouns, and that the underlying case can be any case between the 2nd (accusative) and the 7th (locative). I punctuate them with a hyphen, following Coulson’s representation of dependant determinative compounds. An example for the 6th case (or genitive), which is probably the most frequent, is *yānaik kōtu* (elephant^N-tusk^N) “the elephant’s tusk, elephant tusk”.

2nd-degree expressions: {(A⁰:N₂)-N₁} constructions (*vaṇṇac cinaiccol*)

Before going back to the exploration of the relative importance of U-syntax and M-syntax in the various domains of the CTamil corpus, it seems important to mention a type of construction which is quite frequent in the corpus, and which seems to result from a combination of a *paṇput tokai* with a head noun. That construction is dealt with by TC26c,²¹ and

²¹ *Aṭai ciṇai mutal eṇa murai mūnru um mayānkāmai // naṭai peṇru iyalum vaṇṇac cinaic col*

“Ne pas brouiller l’ordre des trois: (1b) / épithète, membre et principal; (1a) / C’est ainsi qu’ils ont la possibilité de s’employer, (2a) / Les mots à couleur et membre (2b).” (Chevillard, 1996). Although the *sūtra* only mentions “colour” for the adjectival component, other qualities are possible, as seen in the examples provided by the

the standard examples are:

- *ceṅ kāl nārai* ((red^{A-m}:foot^N)-heron^N) “red-footed heron”
- *perun talaic cāttan* ((big^{A-um}:head^N)-Proper_Noun^N) “big-headed Cāttan”

The sūtra describes a situation where 3 components combine in a specific order and become what is called a *vannac cīṇaic col*: inside it, the first element is called *aṭai* “adjunct”, the second is called *cīṇai* “part” and the third is called *mutal* “whole”.

Quantifying the proportions: the notion of U-density

Since I intend to define some tools for comparing the stylistic characteristics of different poems, I shall now introduce some numerical parameters to be used for making these comparisons possible. They are the following:

- **u** is the number of U-components (in this case, $u = 3$)
- **U** is the total number of items in the U-components (in this case, $U = 9$)
- **R** is the number of items in the reduced text (in this case, $R = 6$)
- **T** is the total number of items in the full text (in this case, $T = 12$)

Since the heads of the U-components are counted when determining both R and U, it should be clear that we always have:

$$T = U + R - u$$

Finally, in order to be able to compare texts which do not have the same length, we shall define the U-density and the R-density, in the following way:

- U-density = U/T (in this case, U-density = $9/12 = 75\%$)
- R-density = R/T (in this case, R-density = $6/12 = 50\%$).

The sum of the R-density and of the U-density for a given text is usually greater than 100%: there is some overlapping, due to the fact that the heads of the U-components are ruled by both the U-syntax and the M-syntax.

U-density variations inside the CK corpus.

We now turn back to data exploration and examine the CK corpus. This is the smallest of the corpora considered for this study: an exhaustive study of some of its features is therefore easier. As already stated, it consists of 38 poems containing 4 lines each, and I shall now proceed to expose how the parameters previously defined can be applied to it. Even with a very small corpus like this, it is of course not possible to justify here every step that leads to the results found as was done earlier in the case of the 2-lines sample text. For each poem inside the corpus, the following procedure was applied:

- determine T (not counting the 3 clitic particles $-\bar{e}$, $-\bar{o}$ and $-um$).
- determine U, u and R
- compute the U-density

After making the computation for each poem, the results were ordered from lowest to highest, and the mean value was computed, and found to be close to 35%. The results can be seen in the following chart.

Chart 1: U-density in the CK corpus

| U-Density | Poem Number | Adjectives (number of) | V ⁰ (number of) | Additional remark |
|-----------|-------------|-------------------------------|-------------------------------|--------------------------------|
| 0% | {KT 102} | nil | nil | s = 4 |
| 0% | {KT 120} | nil | nil | s = 3 |
| 0% | {KT 20} | nil | nil | |
| 0% | {KT 399} | nil | nil | |
| 0% | {KT 4} | nil | nil | s = 3 |
| 0% | {KT 97} | nil | nil | s = 4 |
| 15% | {KT 17} | nil | 1 | s = 4 |
| 16% | {KT 42} | 1 | nil | IL feature (<i>nāṭa</i>) |
| 18% | {KT 6} | (1 D item : <i>naṇam</i>) | nil | s = 4 |
| 19% | {KT 44} | <i>naṇam</i> | 1 | s = 3 |
| 19% | {KT 93} | 1 2 | nil | |
| 24% | {KT 137} | 2 | nil | |
| 26% | {KT 124} | 1 | 1 | |
| 26% | {KT 63} | 2 | 1 | |
| 28% | {KT 135} | 1 | 1 | |
| 33% | {KT 33} | 2 | nil | |
| 33% | {KT 96} | 1 | nil | IL feature (<i>nāṭaṅku</i>) |
| 35% | {KT 371} | 0 | 1 | IL feature (<i>nāṭaṅotu</i>) |
| 38% | {KT 166} | 2 | 1 | |

| U-Density | Poem Number | Adjectives (number of) | V ⁰ (number of) | Additional remark |
|-----------|-------------|---------------------------|-------------------------------|--------------------------------|
| 38% | {KT 186} | 2 | 1 | IL feature (<i>nāṭarku</i>) |
| 40% | {KT 157} | 1 (<i>tūu</i> ?) | 2 | |
| 40% | {KT 188} | 2 | 1 | |
| 40% | {KT 37} | 2 | nil | s = 3 |
| 42% | {KT 122} | 3 ? | nil | s = 3 (or 4 ?) |
| 45% | {KT 171} | 2 | nil | |
| 47% | {KT 369} | 1 | 1 | |
| 47% | {KT 3} | 3 | nil | IL feature (<i>nāṭaṇotu</i>) |
| 50% | {KT 227} | 2 | nil | |
| 55% | {KT 315} | 2 | 3 | IL feature (<i>nāṭan</i>) |
| 55% | {KT 39} | 3 | 2 | |
| 55% | {KT 71} | 3 | 2 | ? IL feature |
| 57% | {KT 384} | 3 | 2 | (<i>kāṇavar</i>) |
| 59% | {KT 68} | 3 | 1 | |
| 60% | {KT 47} | 4 | 1 | |
| 62% | {KT 109}} | 2 | 2 | |
| 62% | {KT 116}} | 3 | 2 | IL feature (<i>tuṛaiṇan</i>) |
| 78% | {KT 119} | 4 | 1 | |
| 80% | {KT 1} | 3 | 1 | |

As can be seen on the chart, there is a wide range of value for the U-density parameter, and special attention will be given to the extremes. We can notice for instance that the highest U-density is seen in poem KT1, which is in fact, as noted by E. Wilden, probably the original *Kaṭavul vālttu* ("invocation to God") of the KT anthology, and, as such, probably later than the other poems. On the other hand, we have a set of 6 poems for which the computed U-density is 0%, which means that they do not contain any *tokai*, and for 4 of these 6, I have indicated that the s parameter is greater than 1. This means that they contain more than one sentence, a non-written rule of CTamil poetry (disobeyed in their case) being that a poem should consist of a single sentence. Another feature which has been indicated in this chart is the IL (or "Interior Landscape" feature) which shall be explained later.

Before examining more in detail one of the complex poems, it seems useful to give other statistical information on this small corpus. For instance, the number of adjectives and of *paṇput tokai*-s (or adjectival compounds) which is found in these 38 poems is 62. The distribution is of course irregular because 9 poems do not contain any adjectives whereas some contain 4. These adjectival compounds can themselves be part of more complex constructions like those following the $\{(A^0:N_2)-N_1\}$ template which occurs 14 times in this small corpus. Another possibility is to have a combination of 2 adjectives with one noun, like for instance

in *netu veḷ nilavu* ((long^{A0};white^{A0}): moonlight^N)²² (KT 47).

As far as verbal compounds and extended verbal compounds are concerned, it is to be noted that there are 24 V⁰ forms in this corpus. Since the corpus has 152 lines, this means that the V⁰-density is 16%. This can be compared with the characteristics of longer corpora, like KT and NA, in which we find, respectively, 710 and 1564 V⁰ forms,²³ which gives us V⁰-densities of 28% and 37%. These differences could have several explanations. One important parameter is that the NA corpus consists of poems which have a number of lines comprised between 9 and 12, whereas the poems in the KT corpus have between 4 and 8 lines. Since the CK corpus is simply the subset of the shortest poems in the KT, we could try to state an observation that

“the longer a poem is, the higher its V⁰-density will be”.

This observation might also be continued the following way:

“the longer a poem is, the more it has to draw on the resources of U-syntax, in which V⁰ forms are an essential component”

The building elements of M-Syntax: verbal morphology

Before continuing our exploration of the CTamil corpus, the traditional classification of verbal forms will be briefly summarized here: there is first a dichotomy between forms which are described as (syntactically) complete (*murru*), i.e. the finite verbal forms, and forms which are described as incomplete (*eccam*), and there is even a relatively marginal group of in-between forms, considered as both *murru* and *eccam* (ME).²⁴ As far as *eccam* “incompletes” proper are concerned, they are further subdivided into the subgroup of the *vinai eccam* (VE), ad-verbal participles (or “converbs”), that can be completed by a verb, and the subgroup of the *peyar eccam* (PE), “ad-nominal (or relative) participles”, that can be completed by a noun. Further subdivisions inside PE and VE are made by mentioning the corresponding forms in the paradigm of *ceyral* “to do”. To sum up, we have:

²² This example will be commented on when I examine poem KT47. The use of “;” imitates Coulson’s notation for *dvandvas*.

²³ These figures are based on indices kindly communicated to me by Eva Wilden.

²⁴ Concerning these *murr(u)-eccam* forms, see Steever [1988] p. 45 & p.51.

Chart 2: CTamil verbal paradigm

| Main Groups in the verbal paradigm | Subgroups | Frequent English Equiv. | Statistics for the NE corpus | |
|---|---|--|---|--|
| | | | occur. | Distinct items (vocabulary) |
| PE (<i>peyar eccam</i>) “relative participle” | PE1 : <i>ceyyum</i> PE2 : <i>ceyta</i> | non-past (imperf.) past (perfective) | 86 140 | 70 100 |
| V ⁰ (plain verbal root) | | | 215 | 116 |
| VE (<i>vinai eccam</i>) (several types of converbs) | VE1 : <i>ceytu</i> VE2 : <i>ceyyū/(ceyyā)</i> VE3 : <i>ceypu</i> VE4 : <i>ceyteṇa</i> VE5 : <i>ceyyiyar</i> VE6 : <i>ceyyiya</i> VE7 : <i>ceyiṅ</i> VE8 : <i>ceya</i> | absolute absolute absolute conditional “infinitive” | 200 2 8 4 1 2 15 183 | 129 1 7 4 1 2 13 92 |
| VM (<i>vinai murru</i>) “finite verb” | inflected for PNG and tense | | not computed | not computed |
| ME (<i>murr-eccam</i>) | inflected for PNG and tense | | not computed | not computed |
| VAP (<i>vinaiyāl anaiyum peyar</i>) “participial noun” | inflected for PNG, tense and case. | | not computed | not computed |
| TP (<i>tolir peyar</i>) verbal noun | inflected for case | | not computed | not computed |

Inside this (incomplete) list of forms that together constitute the verbal paradigm it must be noted that the V⁰ type, which we have already mentioned in connection with the “unmarked syntax” is not considered by the CTamil grammarians as a stand-alone form but is mentioned only as a constituent of the *vinait tokai* “verbal compound”.

Syntactic depth: the case of a poem with high U-density: KT 47

We now examine a slightly more complex poem from the CK corpus, which has a U-density of 60%. Although the U-density is a useful parameter for a first exploration, it has to be completed by other parameters. My purpose shall now be to show that the U-syntax elements can not only be frequent in a poem, but that they can also be used, to some extent, recursively, and that the degree of complexity reached

varies from poem to poem.²⁵ That degree notion was already in fact implicit when I wrote that the $\{(A^0:N_2)-N_1\}$ template was used to build second degree expressions. The text of the poem will be directly given with grammatical tags and punctuation, without going through the stage of undoing the sandhi, as was done for the first text sample examined.

[[(((karum^{A-um}:kāl^N)-vēnkai^N)-vī^N)-uku^{V0}):turukal^N]
 [(irum^{A-um}:puli^N)-kuruḷaiyiṅ^{N-obl}] tōṅrum^{V-pe1} kāṭṭu^{N-obl} itai^{Postp}.
 elli^N 26 varunar^{N-vap} kaḷaviṅku^{N-dat}.
 nallai^{A-2nd-sing} allai^{neg.2nd-sing} [(neṭu^{A0};veḷ^{A0}):nilavu^N] ē^{voc}.

[[(((black^{A-um}:foot^N)-Tree(spec.)^N)-flower^N)-fall^{V0}):ROCK^N]
 [(big^{A-um}:tiger^N)-CUB^{N-5th.case}] appear^{V-pe1} forest^{N-obl} through^{Postp}.
 night^N come^{V-agent} secret^{N-dat}.
 good^{A-2nd-sing} not^{neg.2nd-sing} [(long^{A0};white^{A0}):MOONLIGHT^N] ē^{voc}.

The first line in the poem is the description of a rock and boils down to one item in the reduced sentence. The noun "rock" is the head of an extended verbal compound which has for its other argument a case compound. The first component of that case compound is itself a 2nd-degree expression. This makes the line into a fifth-degree U-expression, that can be translated as:

"ROCK(S), on which fall the flower(s) of the black-footed *Vēnkai* [tree]"

The second line contains a 2nd-degree expression which has its head in the 5th case (the case used for comparisons):

"like [the] CUB(S) of [a] big tiger"

The last line contains the U-expression "long white MOONLIGHT" which

²⁵ The hope that it should be possible to quantify precisely the syntactic complexity of CTamil poems, and that this might be useful for clarifying their chronology has been entertained by me ever since I read in Renou[1956] the following, about Sanskrit compounds: "comment la langue est-elle passée d'un état ancien – celui que reflète le Rgveda – où les composés étaient brefs et relativement peu nombreux, à l'état que l'on connaît par la littérature classique, où le procédé a pris un développement que J. Bloch qualifiait justement de « monstrueux » (*Indo-aryen*, p. 106)?"

²⁶ The status of *elli* is not absolutely clear. Instead of seeing it as a noun (as was tentatively done here), it could be considered as the *viṅai eccam* (VE1) of an impersonal verb *ellutal* "to-be-night, to-become-dark". In that case, *elli varumar* would mean "those who come when it is night/when it becomes dark". CTamil is not a well-understood language!

was already mentioned and which must have appeared as striking because the poet's real name was forgotten and he is only mentioned in the anthology as "He-of-the-long-white-moonlight-[formula]".²⁷

The reduced sentence is:

ROCK^N CUB^{N-5th.case.} appear^{V-pe1} forest^{N-obl} through^{Postp.}
 night^N come^{V-agent} secret^{N-dat.}
 good^{A-2nd-sing.} not^{neg.2nd-sing.} MOONLIGHT^N ē^{voc}

That sentence itself starts with the description of a wilderness/forest. The noun "forest" is the head of a relative participle which is also connected with 2 more arguments, which happen to be the heads of the first 2 U-expressions we have just described. This gives us:

"Through [the] forest where ROCK(S) appear like CUB(S)"

The successive embeddings strategy which was seen in the U-components is also seen in the reduced sentence, although it is under the rule of the M-syntax. This is a consequence of the (generally observed) constraint that poems should consist of one single sentence. One could, as a consequence, try to define a degree of embedding for reduced sentences also. A full translation of the reduced sentence is easier to read in English if all the components are inverted:

"Ô MOONLIGHT! you are not good, for the secrecy, of those who come at night, through [the] forest where ROCK(S) appear like CUB(S)."

A full translation of the poem is obtained by replacing the 3 words in small capitals by the translations of the corresponding U-expressions.

Interior Landscape (IL) feature: use of a pivotal *utaipeyar*

While examining the U-density of the 38 poems in the CK corpus, it was stated that 7 (or 8) of them had a special feature which I designated as the "Interior Landscape" (IL) feature. What I was referring to is the fact that in these poems, we find a word in a syntactically pivotal situation, connecting two apparently independant discourses, although one is often reinforcing the other, in an allusive way (through some *ullurai uvamai*). The first discourse is a description of a landscape and the second one is the description of a love situation. The pivotal word syntactically

²⁷ "Long white moon" probably means "white moon that shines for too long".

connects the two. Typical examples are *nāṭan*, *tuṛai*, *ūran*, *verpan*, *cērpṇan*. CTamil grammatical commentators consider these terms as *uṭaip peyar*-s “nouns of possession” and they analyze them as combining a place noun (*nāṭu* “country”, *tuṛai* “ghat, harbour”, *ūr* “town”, *verpu* “mountain”, *cērpṇu* “shore”) with a suffix *-an* which brings the meaning “He who possesses that place” (cf. TC114c & TC165c).

As a typical example, I shall present a simplified extract from KT186

... *mullai mel koṭi ... mukaiyum nāṭarku-t tuyil tuṛantaṇa ... em kaṇ...*
 (mullai^N-(mel^{A0}: koṭi^N)) mukaiyum^{V-pe1} nāṭarku^{N-dat} tuyil^N tuṛantaṇa^{Vvm3PInt}
 em^{N-obl} kaṇ^N
 (Jasmine^N-(soft^{A0}: creeper^N)) bud^{V-pe1} country-He^{N-dat} sleep^N
 renounce^{Vvm3PInt} eye^N

“For Him, of the country where the soft jasmine creeper is budding, my eyes have renounced sleep.”

Seen from the point of view of the interior (or embedded) landscape description, the most important element in *nāṭan* “He-of-the-country” is *nāṭ(u)* “country”. Seen from the point of view of the main sentence, the head, however is the pronominal ending *-an*. The landscape description is here, as is very often the case, anchored onto *nāṭ(u)* thanks to *mukaiyum* which is a relative participle, which acts as the main predicate for the embedded sentence and which has two main arguments: *koṭi* and *nāṭu*. The kernel of the embedded description is thus:

koṭi^N mukaiyum^{V-pe1} nāṭu^N
creeper^N bud^{V-pe1} land^N

“country where a creeper is budding.”

Statistics on the distribution of relative participles and verbal roots

One of the main obstacles to the correct primary²⁸ understanding of a CTamil poem, provided one is familiar with its vocabulary, is to determine how the words are syntactically connected. As far as the U-components are concerned, the absence of marks can be compensated by

²⁸ We shall not be concerned here with understanding the allusions, but shall only be concerned with understanding the syntax, that *axis mundi*.

a familiarity with a stock of formulae. As far as M-syntax is concerned, the marks which are found on a number of items (but not on all) clear the way for construing the reduced sentence, but there are cases of long-distance relationships between items, which may not appear at the first glance.

In this respect, since the relative participles (V^{PE1} and V^{PE2}) and the verbal roots (V^0) play the pivotal role of a verbal kernel connecting several arguments, it has seemed useful to give here some statistics on the most frequent roles that these arguments play with respect to the kernel. Additionally, since the use of V^0 form is characteristic of U-syntax, whereas the use of V^{PE1} and V^{PE2} forms is typical of M-syntax, this exploration, which remains of course a preliminary one, should throw some light on the global economy of the system.

We first start with some statistical figures concerning the total number of such forms in the two biggest corpora that were examined for this study. These figures, which are collected in the following 2 charts,²⁹ are mostly based on the extensive data collected by Eva Wilden (forthcoming publication) for her critical editions (and translations) of two CTamil anthologies, the *Kuruntokai* (KT) and the *Narriṇai* (NA) totalling respectively 4180 and 2504 lines, in most editions.³⁰

**Chart 3a: distribution of V^0 , V^{PE1} and V^{PE2}
inside *Narriṇai* (400 poems)**

| NA (4180 lines) (9 to 12 lines poems) | V^0 (verbal root) | V^{PE1} (<i>ceyyum</i> "which does") | V^{PE2} (<i>ceyta</i> "which did") |
|---|---------------------|--|--|
| occurrences | 1564 | 583 | 853 |
| vocabulary (distinct items) | 343 | 261 | 361 |
| Density | 37.4 % | 13.9 % | 20.5 % |

²⁹ The figures contained in these charts should not be considered as precise to the last digit, because these texts abound in variants. Moreover, and more importantly, the figures given for V^0 are higher than what should be given for V^0N constructions only, because the occurrences of V^0V^{aux} serial constructions have not been factored out. Still, it seemed useful to give them because what matters mostly in them at this stage is their order of magnitude. The presence of V^0 does remain a massive fact.

³⁰ Because of variants, these numbers should be taken as indicative only. The KT contains 399 poems with lengths comprised between 4 and 8 lines (and 2 exceptional 9-lines poems) and the NA contains 400 poems with lengths comprised between 9 and 12 lines (plus seven "exceptional" 13-lines poems).

Chart 3b: distribution of V^0 , V^{PE1} and V^{PE2}
inside *Kuruntokai* (401 poems)

| KT (2504 lines) (4 to 8 lines poems) | V^0 (verbal root) | V^{PE1} (<i>ceyym</i> “which does”) | V^{PE2} (<i>ceyta</i> “which did”) |
|--|---------------------|---|--|
| occurrences | 710 | 361 | 423 |
| vocabulary (distinct items) | 227 | 193 | 225 |
| density | 28% | 14.4 % | 17.5% |

The first bit of information which can be gathered from these two charts is that the use of V^0 forms is a massive phenomenon in CTamil poetry. This is a testimony to the pervasiveness of U-syntax. The massive presence of V^0 forms is all the more impressive since they do not seem to be used in the other varieties of Tamil with which I am familiar (Modern Tamil, spoken or written, H or L, medieval prose Tamil).

A CTamil grammarian's position on argument roles

We shall now take a closer look at the $\{(N_3)... N_2)V^{PE}N_1\}$ and at the $\{(N_2)V^0N_1\}$ syntactic templates which have been presented at the beginning of this exploration and with which we have met quite often. We shall tentatively do it this time from a semantic role point of view, which will be adapted from a CTamil grammarian's conception which is presented several times in the T and illustrated by the commentators. There is indeed a canonical list, first mentioned in TC112c, which consists of 6 prerequisite for an activity to take place; they are the following: P1 *nilan* “place”, P2 *poruḷ* “object (patient)”, P3 *kālam* “time”, P4 *karuvi* “instrument”, P5 *viṇai-mutal* “agent”, P6 *viṇai* “action”. This list resembles the Pāṇinian *kāraka* list but is not identical with it. It appears a second time in TC234c,³¹ in connection with the *peyar eccam* characterization. The commentators enumerate on that occasion two series, respectively, of $\{V^{PE1}N_1\}$ and $\{V^{PE2}N_1\}$, where the N_1 element takes successively the six values. The examples are:

- P1 : *vālum il* “the house where one lives”
- P2 : *kaṟkum nūl* “the book that one studies”
- P3 : *tuyilum kālam* “the time when one sleeps”
- P4 : *vanaiyum kōl* “the stick with which [the potter] fashions [the pot]”

³¹ See Chevillard [1996: 362-365].

- P5 : *ōtum pārppan* “the brahmin who recites”
- P6 : *unnum ūn* “the eating one eats”
- P1 : *pukka il* “the house one has entered”
- P2 : *uṅṭa cōru* “the rice one has eaten”
- P3 : *vanta nāl* “the day one has come”
- P4 : *venra vēl* “the spear with which one has vanquished”
- P5 : *āṭiya kūṭṭan* “the dancer who has danced”
- P6 : *pōyina pōkku* “the going one went”

Among these 6, the last one, P6, is probably slightly mysterious to some readers. What seems to be at stake is the fact that inside a verb expressing some activity, the process itself is accessible to some syntactic transformations like relativization.³²

In the case of V^0 , the T, which does not state anything specific regarding the formation of the *vinait tokai*, could not be expected to give any similar details. However, its commentators do indeed give a list similar to the ones found for V^{PE1} and V^{PE2} . The list is the following:

- P1 : *āṭu aranku* (dance^{V0} stage) “dancing stage”
- P2 : *cey kunru* (make^{V0} hill) “artificial hill”
- P3 : *punar polutu* (unite^{V0} day) “day of uniting”
- P4 : *ari vāl* (cut^{V0} blade) “cutting blade”
- P5 : *kol yānai* (kill^{V0} elephant) “murderous elephant”
- P6 : *cel celavu* (go^{V0} going) “the going one goes”

Statistics on argument roles for V^0 constructions

The universal classification proposed by the CTamil grammarians (for the semantic role of the head of a relative participle or of a verbal root) is of course partly a theoretical one, but since it was proposed by, possibly, native masters of that poetical language, it seems useful to examine how it fits the data which it is supposed to describe. The present examination is based on a subcorpus of the NE corpus, which consists of the 22 first *neytal* poems, totalling 344 lines, in which I could identify 78 V^0 forms, 39 of them being found in $\{V^0N_1\}$ constructions and 39 in $\{N_2V^0N_1\}$ constructions. The global V^0 -density is thus 23%, higher than the one found for the CK corpus. It should however be noted that the distribution

³² When clarifying the meaning of the sixth prerequisite (while commenting on TC112c), the commentator explains that whatever the verb used in a sentence, one can always use the DO verb to ask “what is he DOING?”. The process itself appears thus as the internal object of some DO verb. To eat is to DO the action of eating.

is a very irregular one³³: an extensive examination of a still wider corpus would certainly be called for.

For the 39 $\{V^0N_1\}$ examples, the first impression obtained while examining them one by one is that it seems possible in a number of cases to use the labels provided by CTamil grammarians: for instance a $vel^{V^0}v\bar{e}l^N$ (win^{V0}:spear^N) “victorious spear” (AN60-12 & AN70-13) is indeed a “spear **with which** one (can) win” and can appear as a good P4 example. However, in a greater number of cases the choice seems arbitrary. One can indeed attribute agentivity to the sea (*kaṭal*) and consider $m\bar{u}l\bar{a}nku\ k\bar{a}ṭal$ (roar^{V0}:sea^N) “roaring sea” (AN220-12) as an example of P5. But what about $\bar{o}nku\ c\bar{i}nai$ (rise-high^{V0}:branch^N) “high-rising branch(es)” (AN20-5) and $paṭu\ c\bar{i}nai$ (lean-downwards^{V0}:branch^N) “lowering branch(es)” (AN210-12)? It seems arbitrary to categorize the branch as an agent or as a patient. A more promising direction in that particular case might be to examine these 2 examples in their wider context, where the tree to which the branches belong is mentioned.³⁴ And we could ask similar questions for the greater part of the 39 examples. It should be added that the main difficulty in attributing the labels (P1 to P6) to the head noun is of course the minimality of the syntagm. It is somehow easier to attribute semantic roles when there is more than one argument, as will be seen when we examine the $\{N_2V^0N_1\}$ examples. When the only argument for the V^0 is the head noun, and especially if the phenomenon is seen through the lenses of a language like English (or French), the impression obtained, and probably useful from a pedagogical point of view, is that CTamil only has a small stock of *real* adjectives, but it can use plain verbal roots as quasi-adjectives. These verbal roots no longer refer to actual events or to actual descriptions (in as much as what appears in a poem can be said to be actual!), but they then express aptitudes, permanent states, etc.

Turning now to the 39 $\{N_2V^0N_1\}$ examples, it seems somehow easier to attribute semantic role labels to the two arguments N_1 and N_2 . However, it often still seems impossible to be very precise in assigning the labels, especially when we try to distinguish between a P5 (“agent”) and a P2 (“object, patient”). A more general label like “participant in the process”, which could also include P4 (“instrument”) seems easier to use. Grouping together P1 (“place”) and P3 (“time”) under a more general “Locator” label, and renaming P6 (“action”) as “Process”, I have

³³ For instance, the global number of V^0 forms oscillates between 0 (for AN10) and 8 (for AN190).

³⁴ The syntactic treatment of whole-part relationships is a complex area. One interesting feature in this field is the existence of double subject constructions, which seem to have already been a topic of concern for CTamil grammarians themselves (cf. for instance TC61c in Chevillard [1996: 137-139]).

tentatively assigned 25 among the 39 examples to the squares of a double entry chart, to be found below.

Chart 4: distribution of $N_2V^0N_1$ syntagms (in the NE corpus)

| $\downarrow N_2(V)N_1 \rightarrow$ | Locator | Participant | Process | Total |
|------------------------------------|---------|-------------|---------|-------|
| Locator | | 5 | | 5 |
| Participant | 12 | 6 | 2 | 20 |
| Total | 12 | 11 | 2 | 25 |

As already emphasized, what makes it easier to attribute significant semantic roles to the arguments in $\{N_2V^0N_1\}$ examples, is the fact that there are more than one. This can first be illustrated by the examples where I have labelled one of the arguments as "Locator" and the other one as "Participant". This gives us 2 possible types: $\{N^{Participant}V^0N^{Locator}\}$ and $\{N^{Locator}V^0N^{Participant}\}$. The first one is illustrated by examples like:

- *pū nāru parappin* (flower^N smell^{V0} expanse^{N-obl}) (AN 130-8) "in the vast stretch perfumed by flowers"
- *tirai payil aḷuvam* (wave^N be-frequent^{V0} sea-expanse^N) (AN 210-5) "the sea expanse where there are many waves"
- *curumpu ār kaṇṇi* (bee^N be-full^{V0} wreath^N) (AN 180-6) "the garland on which bees were humming"

The second one is illustrated by examples such as:

- *vicumpu aṇi villin* (sky^N adorn^{V0} bow^{N-obl}) (AN 210-4) "like a rainbow"
- *viḷavu aṇi makalir* (festival^N adorn^{V0} women^N) (AN 70-12) "women adorning themselves at time of festivals"

In the cases where both arguments are labelled as "Participants", i.e. $\{N^{Participant}V^0N^{Participant}\}$, we can observe a polarity between an agent and a patient, as in:

- *koṇṭal iṭu maṇal* (East-Wind^N place^{V0} sand^N) (AN 20-7) "Sands heaped by the East Wind"
- *viḷi aṇi ṇamali* (call^N recognize^{V0} dog^N) (AN 140-9) "dogs that recognize (familiar) voices [and bark on strangers]"

The label "Process" appears in several configurations. One possibility is seen in the example [*tunai puṇar uvakai*]-y-ar (AN 30-3), which should be analysed in two stages. Firstly, a $\{N^{Participant}V^0N^{Process}\}$ element is formed:

- *tunai puṇar uvakai* (companion^N unite^{V0} rejoicing^N) "rejoicing in

the uniting with companions”³⁵

Secondly, a pronominal suffix is added to the extended compound giving as a meaning: “those who rejoice in uniting with companions”.

There are however types of examples where the Locator/Participant/Process grid does not seem appropriate for characterizing the N_1 and N_2 arguments. Among those, 2 deserve special mention. The first one is the case where the 2 arguments are in a relation of meronymy (see Chart 5a). The second is the case the 2 arguments are in a relation of comparison (see Chart 5b).

Chart 5a: distribution of $N_2V^0N_1$ syntagms in the NE corpus (whole-part relationship)

| $\downarrow N_2(V^0)N_1 \rightarrow$ | Part (meronym) | Whole (holonym) | Total |
|--------------------------------------|----------------|-----------------|-------|
| Part (meronym) | | 2 | 2 |
| Whole (holonym) | 2 | | 2 |
| Total | 2 | 2 | 4 |

As a comment to Chart 5a, the $\{N^{\text{Whole}}V^0N^{\text{Part}}\}$ template can be illustrated by

tālai vīl kayirru (fragrant_screwpine^N fall^{V0} rope^N) (AN 20-6) “aerial roots of the fragrant screwpine” (VMS translation)

The $\{N^{\text{Part}}V^0N^{\text{Whole}}\}$ template can be illustrated by
talai avil tālai (sprout^N release^{V0} fragrant_screwpine^N) (AN 90-3)
 “fragrant screwpine of loosening sprouts”

³⁵ As an additional argument for explaining that this type of analysis is not an invention by me but is in line with a traditional analysis by a grammatical commentator, let me quote a similar example drawn from the *Kalittokai* anthology and given by Cēnāvaraiyar under TC234c: “Lorsqu’il est dit: *niṅ-mukaṅ kāṅu-maruntinē n-ennum-āl* «s’il dit: J’ai pour remède [que] de voir ton visage» (*Kali* 60_19) étant donné que [le poète] appelle «remède» (*maruntu*) le fait de voir (*kāṅci*), l’expression *kāṅu-maruntu* «remède [que] de voir» a pris [pour sa construction] le nom de l’action (*viṅaip-peyar*).” (Chevillard [1996, p. 364].

Chart 5b: distribution of $N_2V^0N_1$ syntagms in the NE corpus (comparisons)

| $\downarrow N_2(V^0)N_1 \rightarrow$ | <i>upamāna</i> (comparable object) | <i>upamēya</i> (object compared) | Total |
|--------------------------------------|---------------------------------------|-------------------------------------|-------|
| <i>upamāna</i> | | 2 | 2 |
| <i>upamēya</i> | 2 | | |
| Total | 2 | 2 | 4 |

As a brief comment on chart 5b, I shall give only one example:

kaṅ pōl neyṭal (eye^N resemble^{V0} Neyṭal-[flower]^N) (AN 170-4) “eye-like Neyṭal flower”

The Semantic spectrum of a Syntactic template

As a global comment on this tentative classification of $\{N_2V^0N_1\}$ constructions, it is noticeable that there is a clear dominance of some role sets for the N_1 and N_2 positions. That role set will be referred to as the semantic spectrum of the $\{N_2V^0N_1\}$ construction. According to our statistics for this short sample, the most prototypical $N_2V^0N_1$ U-component is of the type $\{N^{\text{Participant}}V^0N^{\text{Location}}\}$ (12 items). This type of U-component is more frequent than the $\{N^{\text{Participant}}V^0N^{\text{Participant}}\}$ type (6 items), which itself is more frequent than the $\{N^{\text{Location}}V^0N^{\text{Participant}}\}$ type (5 items). Of course, these figures are too small. The spectral analysis that these tentative results provide has to be verified in the future against those for larger corpora. It could be useful for scrutinizing the homogeneity of corpora. In the following section, I shall try to compare very briefly this spectrum with those for $N_2V^{\text{PE1}}N_1$ and $N_2V^{\text{PE2}}N_1$ constructions (and for larger ones). It will be only a brief sketch because such constructions can have a larger number of arguments and can span several lines each. A full description of M-syntax cannot be short.

Statistics sketch for argument roles in $\{\dots N_3\} \dots N_2 \dots V^{\text{PE}}N_1\}$ constructions

I shall now briefly present the results of two preliminary explorations done on the NE corpus, targeting the V^{PE2} and the V^{PE1} occurrences. My first remark is that whereas in $N_2V^0N_1$ items there is a balance in weight between N_2 on the left and the head noun N_1 on the right, the situation is usually much more dissymmetrical between the two sides of a V^{PE2} or of a V^{PE1} : combined with what stands on its left, the V^{PE} is the last element in a relative clause which can be quite long and which draws on most of

the resources of M-syntax. For that reason, it is difficult to fit the various possibilities inside the squares of a double entry chart (as was done in charts 4, 5a and 5b). It often seems possible to recognize semantic roles for the noun arguments, but the establishment of the spectrum would require the use of charts with a number of dimensions greater than 2. Even if possible, this would exceed what can be accomplished here. Moreover, because of the dissymmetry between the two sides of a V^{PE} , the priority seems to be to establish the statistics for the head noun N_1 itself. This is what was done on a subset of the NE corpus, and the results are as follows:

- For the V^{PE1} (i.e. non-past) relative participle, the head noun is more frequently a locator than a participant.
- For the V^{PE2} (i.e. past) relative participle, the head noun is more frequently a participant than a locator.
- Examples where the head noun can be categorized as “Process” (or P6) seem to be quite rare but at least one example was found with a V^{PE1} (in KT37).

In order to illustrate the complex way in which all these forms are used, it has seemed to me that the best method would be to give a “real-life” example, which shows most of them. This is what the next section will be devoted to.

Multi-level Syntactic analysis of a 14-lines poem (AN 320)

The poem chosen is a *neytal* (sea-shore setting) poem, AN320, from the *Akanānūru* anthology, and I shall first give its complete syntactical analysis, without however going through the preliminary stage of undoing the sandhi. Because of the high level of intricacy brought about by the successive embeddings, it has appeared necessary to use distinct sets of paired delimiters, in the following way:

- “{“ and “}” are used for level 0 of M-syntax
- “[“ and “]” are used for level 1 clauses in M-syntax
- “<” and “>” are used for level 2 clauses in M-syntax
- “{“{“ and “}”}” are used for level 3 clauses in M-syntax
- “(“ and “)” are used for U-syntax

Inside a zone/clause delimited by a pair of delimiters, the nominal arguments are distinguished by numerical indices as before, but these indices are flanked by a lower case letter which is a unique identifier for

the delimited zone. For instance N_{4b} , alias *mīn* "fish", is the 4th argument inside "clause b" which describes its street-selling. In case a noun is the head of a relative participle and is a nominal argument for another participle, it will as a consequence be attributed two indices. The same *mīn* "fish" can also be referred to as N_{1a} in the clause that describes its capture by fishermen.

- { [<(ōnku^{V0}:tirai^N)-parappin^{N-obl}>_{N(3a)}
 <(vānku^{V0}:vicai^N)-kolī^{V-VE1}>_(a)
 <timilōn^N>_{N(2a)}
 <tanta^{V-PE2}>_(a)
 <(kaṭum^{A-um}:kaṇ^N)-(vaya^D:mīn^N)>_{N(1a)}] _{N(4b)}
 *[[(talai^N-aṇi^{V0}):alku^N]- (celva^D:taṅkaiyar^N)]_{N(3b)}
 [(viḷavu^N-ayar^{V0}):marukin^{N-obl}]_{N(2b)}
 [vilai^N eṇa^{V-VE3}]
 [pakarum^{V-PE1}]_(b)
 [[[kāṇal^N am^{incr.} cirukuṭi^N (peru^{A0}:nīr^N cērppu^N)]_{N(1b)}-a^{Pro-N-voc}]_(c)
 { [(((((malar^{N-ēr}^{V0}):(uṇ^{V0}:kaṇ^N)) -(em^{N-obl}-tōlī^N)) -evvam^N]-
 alar^{V0}):vāy^N]_{N(1d1)}
 [nīnka^{V-VE3}]_(d1)
 [nī^N arulāy^{V-ME-neg-2nd.Sg} poyppinum^{V-concessive}]_(d2) }_(c)
 { [< { { (neṭum^{A-um}:kaḷi^N) } }_{N(2e)} { {tuḷaiiya^{V-PE2}} }_(e) { { (kurum^{A-um}:kāl^N)-
 annam^N } }_{N(1e)}] >_{N(4f)}
 <(aṭumpu^N-amar^{V0}):ekkar>_{N(3f)}
 <(am^D:cīrai)>_{N(2f)}
 <uḷarum^{V-PE1}>_(f)
 <(((taṭavu^D:nilai^N)-punnai)-tātu^N)-aṇi^{V0}):(perum^{A-um}:
 turai^N)>_{N(1f)}]_{N(3g)}
 [< { { (naṭunku^{V0}:ayir^N) } }_{N(2h)} { {pōlnta^{V-PE2}} }_(h) { { (koṭuñci^N-
 (neṭum^{A-um}:tēr^N)) } }_{N(1h)}] >_{N(2i)}
 <(vaṇtal^N-pāvai^N)>_{N(1i)}
 <citaiya^{V-VE3}>_(i)]_(g)
 [vantu^{V-VE1}]_(g)
 [nī^N]_{N(2g)}
 [tōl^N putitu^{A-VE1} uṇṭa^{V-PE2}]_(g)
 [ñānrai^N]_{N(1g)} }_{N(e)}
 {cūl^N-um }_(c)
 {poy^N-ō }_(c)

{kaṭal^N (aṟi^{V0} :kari)^N -ē}(c)

This poem consists of a single main sentence, but several sub-clauses are embedded in the main clause, and a short description can be made in the following way:

- Clause (a) describes the capture of a fish (*mīn*) [head-noun for (a)]
- It is embedded in clause (b) describing the selling of the fish, that takes place in the coastal area (*cēṟppu*) [head-noun for (b)]
- The word (*cēṟppu*) is the base, in an interior landscape construction, for a pivotal word *cēṟppa* (Vocative “Ô, Lord of the sea-shore”), addressed to the hero, which belongs to the main clause (c).
- The embedded clauses (d1) and (d2) are part of the discourse addressed to him
- Clause (f) is a second landscape description with head-noun *turai* (“ghat”) and it contains an embedded clause (e) describing the activity of a bird.
- That ghat (*turai*) is also the locator in a clause (g) describing, in a flash-back, how the hero took the virginity of the heroine, and containing an embedded clause (i) describing the arrival of his chariot
- The head-noun for clause (g) is a temporal locator in the main clause (c)
- The remaining words are the end of the discourse addressed to the hero.

I conclude this section with a word-for-word gloss and a translation of the poem:

{ [<(rise_high^{V0}:wave^N)-expanse^{N-obl}>_{N(3a)}
 <(bend^{V0}:speed^N)-cause_to_take^{V-VE1}> (a)
 <boat_man^N>_{N(2a)}
 <give^{V-PE2}> (a)
 <(fierce^{A-um}:eye^N)-(strong^D:fish)^N>_{N(1a)]}_{N(4b)}
 [((leaf^N-adorn^{V0}):hip^N)-(pet^D:younger_sisters^N)]_{N(3b)}
 [(festival^N-celebrate^{V0}):street^{N-obl}]_{N(2b)}
 [price^N Quotative^{V-VE8}]
 [announce^{V-PE1}]_(b)
 [((grove^N(adj.suffix)^{incr.} hamlet^N)(big^{A0}:water^N) sea_shore^N]_{N(1b)}-
 (pron.suffix)_{Pro-N-voc} } (c)

{ [(((((flower^N-resemble^{V0}):(be^{V0}:eye^N))-(we^{N-obl}-female_friend^N))
 -suffering^N-gossip^{V0}:mouth^N]_{N(1d1)}
 [avoid^{V-VE8}]_(d1)
 [you^N grace^{V-ME-neg-2nd.Sg.} deny^{V-concessive}]_(d2) }_(c)
 { [< { {(long^{A-um}:back_water)^N } }_{N(2e)} { {stir^{V-PE2} } }_(e) { {(short^{A-um}:leg^N)-
 swan^N } }_{N(1e)} >_{N(4f)}
 <(Veg.Spec.^N-reside^{V0}): sand_dune>_{N(3f)}
 <(beauty^D: wing)>_{N(2f)}
 <adjust^{V-PE1}>_(f)
 <(((curvedness^D:stand^N) - Veg.Spec.-) pollen^N)- adorn^{V0}):
 (big^{A-um}: ghat^N)>_{N(1f)}]_{N(3g)}
 [< { {(tremble^{V0}:sand^N)} }_{N(2h)} { {cut^{V-PE2} } }_(h) { {(Ornament_Sp.^N-
 (long^{A-um}:chariot^N)} }_{N(1h)} >_{N(2i)}
 <(Game_Sp.^N-doll^N)>_{N(1i)}
 <destroy^{V-VE8}>_(i)]_(g)
 [come^{V-VE1}]_(g)
 [you^N]_{N(2g)}
 [shoulder^N new^{A-VE1} eat^{V-PE2}]_(g)
 [day^N]_{N(1g)} }_{N(c)}
 {oath^N-Coord.part.}_(c)
 {lie^N-Interrog.part.}_(c)
 {sea^N(know^{V0}:witness)^N-Emphat.part.}_(c)

"Chief of the maritime tract having small hamlets surrounded by sea-shore groves where the petted sisters, who wear on their waist a garment of leaves, sell in the streets where festivals are celebrated the ferocious and strong fish caught [by their brothers] in their boats, dragging the fishing nets with great force in the expanse of the rising waves in the ocean! Even if you become unfaithful to your word by not bestowing your grace on my friend, who has eyes that are as beautiful as flowers and that are taking in collyrium, to remove her distress and the scandal of the village folk,

Is the oath that you swore on that day when you had first union with her, coming in your tall chariot having a part resembling lotus bud, cutting the loose fine sand so as to destroy the doll made of sand during play, in the big ghat which was made beautiful by the pollen of the *punnai* tree (mast wood tree) of curved stature, where the swan of short legs adjusts its feathers with its beak and dries them (eating the fish) after stirring the long back-water (for fish), sitting on the mound of sand on which the creeper of *atumpu* is resting, also false, for which the sea bore a witness?" (V. M. Subramanya Ayyar)

The limits of the distinction between U-syntax and M-syntax

After explaining for a number of pages that there is a dual syntax in CTamil, I must also now mention several facts that might appear as contradicting the thesis I have been defending. These facts all have to do with the “opacity” (or non-opacity) of the U-components. According to what I have illustrated by a number of examples, the only element in a U-component which should be eligible for syntactic relationships with the elements that belong to what I called the “reduced sentence” should be the head of the U-component. But if that principle holds good, we might possibly have to extend the list of what may be accepted in a U-component. The elements we have considered so far were:

- adjectives (plain root or suffixed form: $A^0/A^{um}/A^m$)
- verbal root (V^0)
- nouns (base form N, but possibly also oblique form N^{obl})³⁶
- apparently uninflected items tagged D (“determiners”)³⁷

In this setup, the only element inside a U-component that can bear a mark is the head noun. However, we might have to extend this list to V^{VE1} forms (the 1st type of converb, see above), because it seems that in some cases, V^0 elements can be specified by a V^{VE1} form, as in the following (simplified) example drawn from KT

((vārtu^{V-VE1}-ilāṅku^{V0}):(vai^{A0}: eyirru^{Nobl})) (KT 14-2)
 ((be_regular^{V-VE1}-shine^{V0}):(sharp^{A0}: tooth^{Nobl})) (KT 14-2)
 “[girl with] sharp teeth, that shine regularly”

Although such examples do not seem to be frequent, we have to define a consistent position in order to deal with them:

- either extend the list of what can be met with in a U-expression (and find out precisely how frequent such cases are).
- or characterize this example as exceptional and find some (possibly chronological) reasons for its being exceptional.

Another difficulty which can be met with if we maintain that there is a

³⁶ A full treatment of that question is outside the scope of this study, because it would require going deeply into the morphophonology (and especially into the question of increments (*cāriyai*)).

³⁷ It should be added that characterizing an item as “uninflected” does not necessarily mean that it can have only one form, because the choice between variant forms can be dependant on phonological conditions, as seems to be the case for an item like *taṭam/taṭavu*.

strict separation between what belongs to U-syntax and what belongs to M-syntax is that in some poems, and especially in long poems (but not exclusively) some element which clearly should not be considered as belonging to a U-component seems to be related to an element which belongs to a U-component but which is not its head. One such example is the following:

..... entai^N tanta^{V-PE2} [(((kolu^{A0} : mīn^N) - uṇaṅka^N) - paṭu^{V0}) : pu^N]
 ḍoppi^{V-VE1}
 our_father^N give^{V-PE2} [(((fat^{A0} : fish^N) - drying_pieces^N) - befall^{V0}) :
 bird^N] chase^{V-VE1}

“chasing the birds, that befall on the pieces [left in the open] for drying, of the fat fish, that our father brought ...” (AN 20, 1-2)

The problem, if we want to consider [(((kolu^{A0} : mīn^N) - uṇaṅka^N) - paṭu^{V0}) : pu^N] as a 4th-degree U-expression, is that what “our father brought” is certainly not *pu!* (“the birds”), and that it is not the *uṇaṅka* either, but that it is most probably *kolu mīn* (the “fat fish”). But that makes an element which is apparently deeply seated inside a U-expression syntactically accessible to an element which is clearly outside it, and some explanation has to be proposed for that. I could tentatively suggest two (which are not necessarily exclusive of each other):

- a metrical one: poems are divided into lines (*aṭi*) and our would-be U-expression sits on one line whereas *entai tanta* sits on the former line.
- a chronological one: our poem belongs to a later period, where the duality between the two syntaxes was becoming blurred.

These two cases cannot of course be solved without a much more thorough exploration of a corpus which contains (limiting ourselves to KT, NA and AN) almost 14000 lines, but the questions have to be kept in mind.

Conclusion and open questions

We have now reached a stage where it is necessary to characterize in a general perspective what has been attempted here. We have in front of us a corpus of poetical texts, composed in a language which does not have now, and perhaps never had, native speakers, although those who

composed these (often complex) texts had a mother tongue which they designated by the same word as the language of these texts: *tamiḷ* "Tamil". We also have access to some grammatical treatises where a distinction is made between *vaḷakku* "(ordinary) usage" and *ceyyuḷ* "poetry", although many consider that what *vaḷakku* refers to is not completely clear. Moreover, as was already stated at the beginning, very little is known with certitude regarding the dates when the poems were composed and regarding the date when they were collected into anthologies. In this context, the distinction I have made between U-syntax and M-syntax is a hypothesis, waiting for confirmation or for falsification. But that hypothesis has the advantage of being rooted in a traditional CTamil grammarians' conception, which is, given the circumstances, probably the closest we can get to a native speaker's intuition.

Among the features that have been examined here, the most mysterious one is probably the use of the V^0 elements (or plain verbal roots) in the verbal compounds (extended or not). The striking fact is that it had a massive presence in CTamil texts (as attested by our statistics) but that it is not present in modern Tamil, and is not seen in medieval Tamil prose either, but seems to be a linguistic element specially cultivated in poetry. That being the case, what seems strange from a descriptive linguist's point of view is the contention of traditional CTamil grammatical commentators that $\{V^0N\}$ constructions are equivalent both to $\{V^{PE1}N\}$ and to $\{V^{PE2}N\}$. Their standard example is *kol-yānai* "murderous elephant", which they declare as being equivalent both to *kollum yānai* "elephant which kills" and to *konra yānai* "elephant which killed", where *kol*, *kollum* and *konra* are respectively the V^0 , V^{PE1} and V^{PE2} of the verb *kollutal*. What seems strange in the equivalence statement³⁸ is that it concerns forms which have a massive presence, both in terms of number of occurrences and of vocabulary richness. A possible explanation is that this equivalence statement implicitly recognizes the duality between the marked syntax and the unmarked syntax. If these two realms are considered as not overlapping,³⁹ then the equivalence statement takes on a different signification, that of a correspondence between the dual components.

Every study brings with itself the after-thought realization of what could have made it more complete and what could be a future

³⁸ A possible explanation might be that they were looking for (far-fetched) explanations of the cryptic *Tolkāppiyam* statement concerning the V^0N : *viṇaiyiṇ tokuti kālattu iyalum* (TC409i).

³⁹ Otherwise, if we see U-syntax and M-syntax as one single system, we have a 3 term opposition between $V0$ (generic /descriptive), $PE1$ (imperfective) and $PE2$ (perfective).

continuation. In the case of the present study, if we consider its emphasis on the examination of V⁰N constructions, it must be admitted that a necessary complement would be the study of V⁰V^{Aux} constructions in the same corpora.⁴⁰ This could even throw some light on our understanding of some V⁰N examples, when N is verbal noun.⁴¹ And this brings us back to the consideration of “linguistic plausibility” which was implicit in my statement of “Limitation 4” at the beginning of this study. Is the language of poetry a language? How must linguists handle it? Is the intensive use of V⁰ forms an innovation by Tamil poets? Or is it an archaic trait, that preserves a feature that was once more widespread⁴²? Can poets “invent” a new language? In which condition and under which influence? Drawing on which resources? An answer to these questions cannot be reached here but they have to be asked, nevertheless.

Bibliographical references

- Britto, Francis, 1986, *Diglossia: A study of the theory with application to Tamil*. Washington, D. C. : Georgetown University Press.
- Chevillard, Jean-Luc, 1994, “L’adjectif dans la tradition tamoule”, in B. Colombat (Ed.), *L’Adjectif: Perspectives Historique et Typologique, Histoire Epistémologie Langage, tome 14, fasc. I*, Presses Universitaires de Vincennes, St Denis.
- _____, 1996, *Le Commentaire de Cēṇāvaraiyar sur le Collatikāram du Tolkāppiyam*, Publication 84.1 de l’Institut Français de Pondichéry.
- _____, 2004, “Ideophones in Tamil: a historical perspective on the X-ənal expressives (*olikkurippu āruppaṭai*)” (pp. 407-433), in Chevillard and Wilden, *South Indian horizons: felicitations volume for François Gros on the occasion of his 70th birthday*. Publications du département d’indologie 94, IFP, EFEO, 2004
- Coulson, Michael, 1989, *A critical edition of the Mālatīmādhava*, Originally presented as the editor’s thesis (doctoral—Oxford), revised by Roderick Sinclair, Oxford University Press, Delhi Bombay Calcutta Madras.
- Krishnamurti, Bh. (Ed.), Colin P. Masica, and Anjani K. Sinha (assoc. Ed.), 1986, *South Asian Languages: structure, convergence and diglossia*, Motilal Banarsidass.
- Lehmann, Thomas, 1994, *Grammatik des Alt tamil, unter besonderer Berücksichtigung des Caṅkam-Texte des Dichters Kapilar*, Franz Steiner Verlag, Stuttgart.
- Renou, Louis, 1956, “Sur l’évolution des composés nominaux en Sanskrit”, *Bulletin de la Société de Linguistique* 52, pp. 96-116, Paris [reprinted in pp. 242-262 of Louis Renou, *Choix d’études indiennes*, EFEO, réimpr. N°9, Balbir N. & Pinault G.-J. (ed.), 1997, Paris.

⁴⁰ The most visible examples are probably those where V^{Aux} is *tarutal* “give”, often found with movement verbs (*eḷu-taru, ili-taru*, etc.) which I tend to see as being oriented towards the speaker. Another example is the case where V^{Aux} is *takutal* “to be fit”, as in *kān^{VO} taku^{VO} vaṇappu^N* (see^{VO} be_fit^{VO} beauty^N) “beauty which deserves to be seen/admired” (AN 220-7).

⁴¹ The example [*tuṇai puṇar uvakai*]-y-ar “those who rejoice in uniting with companions” (AN 30-3), already discussed, might then be seen as based in fact on a serial verb construction, combining *puṇartal* “to unite” and *uvattal* “to rejoice”.

⁴² And, of course, what is the situation in other Dravidian languages?

210 *Syntactic Duality in Classical Tamil*

- Steever, Sanford B., *The serial verb formation in the Dravidian Languages*, Motilal Banarsidass, 1988.
- Subramanya Ayyar, V. M., 1975, *Akanāṅṅuru, part I, kaḷirriyānai nirai (1-120), part II, maṇimiṭai pavaḷam (121-300), part III, nittilakkōvai (301-400)*, translated into English by —, Institut Français d'Indologie, Pondichéry, 3 volumes, [unpublished]
- Wilden, Eva, 2002, "Towards an Internal Chronology of Old Tamil Caṅkam Literature Or How to Trace the Laws of a Poetic Universe", *Wiener Zeitschrift Für die Kunde Südasiens*, XLVI (pp. 105-133).
- _____, 2002, *Glossary and Statistics of the Kuṟuntokai*, Habilitation Thesis, University of Hamburg.
- _____, forthcoming, *Critical edition and translation of NA*.

MLBD SERIES IN LINGUISTICS

General Editor
DHANESH JAIN

Volume XVI

**Old and New Perspectives on
South Asian Languages**

Grammar and Semantics

Papers growing out of
**THE FIFTH INTERNATIONAL CONFERENCE
ON SOUTH ASIAN LINGUISTICS**
(ICOSAL-5), held at Moscow, Russia in July 2003

Edited by
COLIN P. MASICA

**MOTILAL BANARSIDASS PUBLISHERS
PRIVATE LIMITED • DELHI**

First Edition : Delhi, 2007

© Colin P. Masica
All Rights Reserved

ISBN: 978-81-208-3208-4

MOTILAL BANARSIDASS

41 U.A. Bungalow Road, Jawahar Nagar, Delhi 110 007
8 Mahalaxmi Chamber, 22 Bhulabhai Desai Road, Mumbai 400 026
203 Royapettah High Road, Mylapore, Chennai 600 004
236, 9th Main III Block, Jayanagar, Bangalore 560 011
Sanas Plaza, 1302 Baji Rao Road, Pune 411 002
8 Camac Street, Kolkata 700 017
Ashok Rajpath, Patna 800 004
Chowk, Varanasi 221 001

PRINTED IN INDIA
BY JAINENDRA PRAKASH JAIN AT SHRI JAINENDRA PRESS,
A-45 NARAINA, PHASE-I, NEW DELHI 110 028
AND PUBLISHED BY NARENDRA PRAKASH JAIN FOR
MOTILAL BANARSIDASS PUBLISHERS PRIVATE LIMITED,
BUNGALOW ROAD, DELHI 110 007

CONTENTS

| | |
|----------------------------------|------|
| <i>Editor's Preface</i> | ix |
| <i>A Note on Transliteration</i> | xi |
| <i>List of Contributors</i> | xiii |

PART I: HISTORICAL AND ETYMOLOGICAL PERSPECTIVES

VIT BUBENIK 3

On the Evolutionary Changes in the Middle and New Indo-Aryan Systems of Case and Adpositions (with special reference to European Romani)

[Sanskrit, Pali Ardhamagadhi, Apabhramsa, Mod. Hindi; + European Romani, cf. w. Pashto, Turkish]

TATIANA ORANSKAIA 24

Words with the Meaning 'Fear' in Indo-Aryan Languages

[various NIA lgs, esp. Hindi, Panjabi, Garhwali, Romani ("Gypsy"); Dardic & Nuristani; cf. w. OIA, other IE, esp. *Proto-Slavic, *PIE, Armenian]

ALEXANDER A. SIGORSKIY 34

Case, Split Nominativity, Split Ergativity, and Split Accusativity in Hindi: A Historical Perspective

[Khari Boli, Braj, Awadhi, Dakkhini, Nepali, Panjabi]

PART II: CROSS-LINGUISTIC AREAL OR TYPOLOGICAL PERSPECTIVES

[Note: Several papers in Parts I and III also involve cross-linguistic perspectives, as noted]

HANS HENRICH HOCK 65

South Asia and Turkic: The Central Asian Connection?

[data from Sanskrit, Gandhari Prakrit, Avestan, Persian, Tamil, Turkish and Old Turkic, Ossetic, Tocharian A & B, Tibetan, Newari, Burushaski, Khotanese, etc.]

- LIUDMILA V. KHOKHLOVA & CHARANJIT SINGH** 91
Resultative Constructions with Overt Agent-Possessor in Hindi-Urdu and Western NIA Languages
[Hindi-Urdu, Panjabi, Gujarati, Marwari + comp. with Russian]
- HSIN-HSIN LIANG & PETER E. HOOK** 109
The Compound Verb in Chinese and Hindi-Urdu and the Plausibility of the Indo-Turanian Linguistic Area
- V. P. LIPEROVSKY** 144
Notes on the Marking of Actants in Braj (in Comparison with Modern Standard Hindi)
- KARUMURI V. SUBBARAO** 153
The Role of Particles and Clitics in Disambiguation in South Asian Languages
[Hindi-Urdu, Panjabi, Telugu + Tibeto-Burman lgs: Manipuri, Ao, Sema, Mizo]

**PART III: OLD AND NEW DESCRIPTIVE PERSPECTIVES
ON PARTICULAR LANGUAGES**

- JEAN-LUC CHEVILLARD** 177
Syntactic Duality in Classical Tamil Poems
- ALICE DAVISON** 211
Hierarchical Structure and Linear Order: Correlative Clauses in Hindi-Urdu
[comparisons with Sanskrit; also with Sinhala, Malayalam]
- COLIN P. MASICA** 246
Postverbal Subjects in Telugu and Other Languages
[comparisons w. Hindi, Turkish, Gujarati, Russian, Chinese, Spanish]
- JOHN PETERSON** 274
Languages Without Nouns and Verbs? An Alternative to Lexical Classes in Kharia
- DMITRI V. SITCHINAVA** 304
Past Markers in Santali: a Typology-Oriented Approach
[ref. also to Bantu, Romance, Uto-Aztecan, Lezgian, Russian, Marathi, English]

| | |
|---|------------|
| KASHI WALI & OMKAR N. KOUL | 323 |
| <i>Case Doubling in the Kashmiri Possessive: Another Look</i> [comparisons w. Old Georgian, Turkish] | |
| BORIS A. ZAKHARYIN | 337 |
| <i>PaaNini's 'Lopa' and Zeroing in Modern Standard Hindi</i> | |
| <i>Index</i> | 349 |