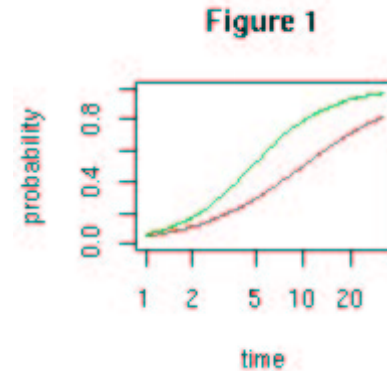


I. Introduction.

The use of statistics to describe language change has already enjoyed notable success in its relatively brief life. While the use of lexicostatistics and glottochronology met with considerable skepticism, more recent studies have returned more unambiguous results. Foremost among these studies have been the ones exploring the so-called Constant Rate Effect (Kroch 1989 and others). This hypothesis seeks to explain the S-shaped curve which best describes how any particular aspect of a grammatical system changes. In this type of curve, change starts slowly, then accelerates, then slows again, as seen in the lines in figure 1. It was previously thought that this curve reflected the underlying mechanism of linguistic change, that the rate of language change itself changed across time.



The Constant Rate Effect does not challenge the description, but instead changes the explanation behind the curve. Instead of the rate of change changing, it is the frequency of the contexts in which a change is found which changes. As a context becomes more frequent, a linguistic change found in or dependent upon that context will also become more frequent. It is also the case that sometime the same change occurs in multiple contexts. Previously it was thought that the rate of change was independent in each context. The Constant Rate Effect shows that this view is false and that the rate of change is constant across all contexts—it is actually the frequency of the contexts which varies. This is one part of the "constant" in the Constant Rate Effect.

The other important part of the CRE which remains constant is the rate of change itself. While it is undeniable that the curves show an apparent change in steepness, under the proper mathematical viewpoint this change disappears. The mathematical construct necessary here is borrowed from biology and called the *logistic*, described by the equation

$$p = \frac{e^{k+st}}{1 + e^{k+st}}$$

where p is the rate (probability) of attestation, t is the historical time, s is the (constant) slope of the curve, and e and k are constants.¹ On the right-hand side of the equation, the only factor changing is t . When t is small, at the beginning of the linguistic change,

¹ E is the mathematical constant, the base of the natural logarithm and approximately 2.718. K describes where the curve intersects the x -axis. As such, it is necessary for graphing the curve but otherwise uninteresting to the linguist.

the denominator of the fraction will dominate and p will be small, reflecting a low rate of attestation. When t is very large, at the end of the linguistic change, the difference between the numerator and the denominator will be very small and p will be close to 1—meaning a nearly 100% probability of attestation. In between the two extremes, the ratio between the numerator and denominator changes very quickly, reflecting the high "steepness" of the middle of the curve. In this way the logistic equation describes the sigmoid curve.

The methodology for demonstrating the CRE depends on having a relatively good-sized corpus from which many examples of some change can be extracted, both in the positive (the change having taken place) and the negative (the change having not taken place). The relationship between these numbers gives the probability of attestation for that change in the text in question. In order to determine a good time dimension, it is necessary that the texts of the corpus extend across some reasonable length of time, usually on the order of two or three centuries. With these methodological constraints in place, the CRE has been demonstrated for Old English (Kroch 1989), Yiddish (Santorini 1993), Ancient Greek (Taylor 1994), and others.

Difficulties arise, however, when one or more of the criteria for the CRE methodology is absent or faulty. Such is the case for Pali. The Pali literature includes nearly 4 million words in the canonical literature alone, spanning four centuries at the very least.² Pali was used at a time of great linguistic change in India, as the elaborate grammatical systems of Old Indic (*eg* Vedic and Sanskrit) were collapsing and the seeds of the vastly typologically different Modern Indic (*eg* Hindi) systems were just germinating. For example, between Old and Modern Indic the verb system lost the medial voice altogether, lost several modes such as the subjunctive and optative, and reconstructed the non-present system from single inflected forms to periphrastic constructions. The noun system lost the dual number and basically the entire case system, moving from a nominative-accusative system to an ergative. Any of these changes are countable and therefore could serve as the basis of a diachronic study. What is missing is the clear division into historical strata. That is, while there are a lot of texts spread across a lengthy period of time, the correlation between any given text and any particular period is usually unknown. This study will attempt to show how a rough chronology can be induced based on a known historical change, focussing on the shift from multiple formations for finite past tenses ("preterites" or "aorists") to only a single formation.

II. Grammatical Background.

Old Indic, as attested in Vedic and Sanskrit, had a wealth of finite past tenses, including a perfect, an aorist, and an imperfect, each based upon a different stem formation. By the time of even the earliest Middle Indic, however, the perfect was largely gone except for a few fossilized remnants, and the imperfect had been folded into the aorists. The aorists

² Including later works such as commentaries and histories vastly increases both the size of the corpus and its chronological range, but these works are omitted from the current study.

throughout this time were the most productive system, and the system itself showed multiple formations. These are traditionally divided into four types.

Type I (Root) Aorists

Generally speaking this is the most unchanged of the paradigms, although it does contain a mixture of forms. Geiger (1956, p 191) presents the following example of the root aorist:

<i>Sg.</i>	<i>Pl.</i>
1 adam̐ JA III.41 ¹⁰	(adariha) JA. II. 71 ⁴ , Vv. 68.4, 5
2 ado (adā) JA. IV. 240 ¹⁴ , V. 161 ¹²	(adattha) JACo. II. 166 ²¹
3 adā Sn. 303, Mhvs. 7.70	adū adum̐

Geiger notes that the first and second persons plural are not inheritances from the Old Indic root aorists, which would be adā ma and adā ta respectively, but rather forms which have migrated into this paradigm from the Type III/s–aorist. Furthermore, several individual roots also innovate a new first person singular root aorist on the analogy of adam̐, such as ahum̐ from bhū (*contra* expected ahūvaṁ) and akaṁ from kṛ (*contra* expected akaraṁ). Likewise bhū shows an unexpected third plural ahū/ahum̐ *contra* the expected ahūvaṁ, again on the analogy of the paradigm for dā. In Pali, only about half–a–dozen or so roots attest the root aorist.

Type II (Thematic) Aorists

The members of this paradigm can be reflexes of either the Old Indic root aorists or the thematic imperfects. Geiger (*ibid*) offers the following example:

<i>Sg.</i>	<i>Pl.</i>
1 agamaṁ Th1. 258	agamāma (agamam̐ha Sn. 349)
2 agamā Sn. 834	agamatha (agamattha)
3 agamā Sn. 408, Mhvs. 5.42	agamum̐ Sn. 290, Mhvs. 4.36

Within this paradigm the singular forms can usually be regarded as inheritances, and likewise the first plural in –āma. The aspiration on the second plural, rather than the expected agamata, can be regarded as having been imported from either the third singular medial (eg ajā yatha ‘originated’ Dpvs 5.40) or from the Type III/s–aorist ending –ttha. The third plural form must be regarded as an innovation based on the general third plural ending in –um seen across the paradigms, of uncertain origin. A couple dozen roots show thematic aorists in Pali.

Type III (s) Aorist

The Types III and IV are the most productive classes of preterites in Pali and most roots show some forms in these paradigms regardless of whether those roots attested s- or is-aorists in Old Indic. Geiger (*op cit*, p 192) offers the following paradigms:

Sg.	Pl.
1 assosiṃTh1.131	assuṃha S.I.157 ¹² , JACo. III.278 ⁶
2 assosi	assuttha D. II.272 ² , S. I.157 ⁹
3 assosi D. I.87 ¹¹ , Sn p 99	assosuṃ D. I.111 ¹⁰ , Vin. I.18 ³⁵
1 akāsiṃ Th2.74, Vv.1.5	akamha JA.III.47 ⁴
2 akāsi Vv.1.3, Th1.120 ⁷	akattha Vv. 84.38, Mhvs. 12.22
3 akāsi JACo. III.188 ²⁴ , DhCo I.39 ⁶	akāsuṃ Mhvs. 31.99 var lec., akāṃsu Sn. 882, JACo. I.262 ⁶

Within this paradigm, the first singular must be regarded as an innovation. The basis for the innovation is fairly straightforward: an extension of the pattern agamā:agamam̐ as in the Type II aorist or adā:adam̐ in the Type I aorist to the assosi base in the Type III. The third plural is also an innovation but of mixed origin. To the stem is added the standard formatives of the aorist marker -s- and the third plural markers -m̐-, from the third plural imperfects among other sources, and -u- apparently from the third plural perfects. The exact order the markers go in seems to be somewhat variable, however, sometimes producing -suṃ and more frequently -ṃsu. A further issue is the question of the grade of the stem. In Old Indic the stem took vṛddhi grade across the active voice forms. Middle Indic lost the ability to make the vṛddhi grade due to monophongization of -āi- and -āu- so the resulting forms in -e- and -o- are not surprising. What is surprising is the normal stem grade in the first and second plural forms. This must be regarded as having spread from the shortening of -ā- in the first and second plural forms for roots such as kṛ above. Because of the phonological necessity of not having a long vowel before two consonants, the guṇa grade was shortened to the normal grade -a-. The phonological motivation need not have applied to forms in -o-, so the shortening was apparently reanalyzed as a morphologization of a previously phonological rule.

Type IIIb aorists

On the basis of pairs of Type II and Type III aorists such as akā/ akāsi and adā/ adāsi the formative of -āsi, as a blend of the thematic and s-aorists, was widely extended as an aorist marker. Otherwise the morphology of this type follows that of the standard Type III. These must all be regarded as innovations. Only eight verbs attest this formation, which is not surprising given that it is the combination of two others. An example paradigm is the following:

	<i>Sg.</i>	<i>Pl.</i>
1	agamāsim	agamāsimha
2	agamāsi	agamāsittha
3	agamāsi	agamāsimsu

Type IV (is) aorists

As with the Type III paradigms, the Type IV aorists are a highly productive category. Geiger gives the following example:

	<i>Sg.</i>	<i>Pl.</i>
1	agamisaṃ agamiṃ Th1.9	agamimha S. I.202 ³³ (verse)
2	agami Sn. 339, JACo. IV. 2 ¹⁷ agami D. II.264 ⁹ , JACo.	agamittha JACo. I.263 ⁴ , DhCo. III.22 ⁷ agamisuṃ, agamimsu JACo. II.416 ²³
3	VI.366 ²⁴	

Other than the modification of the aorist marker itself, the morphology of this paradigm is basically the same as that of the Type III paradigm. An exception is the form *agamisaṃ*; this is one of the few forms which takes the historical first singular ending *-aṃ* rather than the modified Middle Indic *-iṃ*.

Related to the Types III and IV is a third formation of unclear origin, the so-called *e-*aorists. These follow the pattern of the personal endings of the *s-* and *is-*aorists, but use an additional aorist marker of *-e-*. An example paradigm is:

	<i>Sg.</i>	<i>Pl.</i>
1	yojesiṃ	yojesimha
2	yojesi	yojesittha
3	yojesi	yojesimsu

This formation has two possible sources. The first is the use of *-e-* as a stem formative in the present tense, a reflex of the Old Indic verbs in *-ya*, *-yá* and *-aya*. The regular monophongization of Middle Indic rendered the earlier stems such as **yoj-aya-* into *yoj-e-*. Once the *-e-* was present in the stem, it was easily imported into the other tense formations such as the aorist, where it took the productive Type III endings. Alternatively, von Hinüber (1997) takes the *e-*aorist as stemming from a rereading of an optative (itself formed in *-e-* after monophongization) or some other non-aorist in *-e-* as an aorist, citing the line *nikkhama pāpima mā tathāgataṃvihesesi ... mā te ahoṣi ... dukkhāya* (MN I.332.14f) where the form *vihesesi* can be read either as a present or as a prohibitive aorist after the *mā*. (von Hinüber 1977, p 41). Similar cases further eroded the distinction between present tense or ‘optative’ usages of *-e-* forms and aorist usages, and around this ambiguity a whole paradigm of *e-*aorists was built.

While this ‘optative’ usage may have reinforced the change by adding another source of –e– in verb forms, it seems most likely that the primary source of the e–aorist is the monophthongization of –(a)ya– in the present tense stems. As long ago as Bloch (1877) it was recognized that the present tense stem was becoming (or had become) the basis for the entire verb paradigm. Regardless of the origin, the entire e–aorist paradigm must be regarded as innovations.

There is a further variation on the Type IV aorist, brought about when the formative –is– (< OIA –iṣ–) is added to a stem already ending in e, as in a causative stem or present stem in e. This could be resolved either as a e–aorist as above, or by decomposing the final e into ay and adding the expected is–aorist endings to that stem. Indeed, many verbs show both behaviors.

Similar again to this construction are a few cases where the –is–formative is added to a stem which ends in –ā, resulting in an epenthetic –y– to avoid vowel hiatus. In some of these cases there is a corresponding present tense formation in –āy (eg, *gāyati*, *jhāyati*) but in others there is not.

Finally, there are other odds and ends within the past–tense system which must be addressed. There are the occasional forms which are reflexes of Old Indic perfects, most commonly *āha* but also including *asā*, *āhu*, *jagāma*, *babhūva*, *vidū*, and *vidum*, (von Hinüber 1994, p 173) and *papāta* (Fahs 1985, p 300) but these forms are unproductive and do not form any kind of paradigm. There are also traces of reduplicated aorists such as *ajjhapattā* from **adhi-a-paptat* (root *pat* ‘to fall’), but these are always reanalyzed within Pali and fit into one of the other Types, usually the thematic Type II. These formations will play no further role in this study.

Of all these formations, however, only Types III and IV and their variants, collectively known as the sigmatic aorists, are truly productive. This can be seen partly by the fact that every verb which attests an aorist attests one of the sigmatic forms, while the same is not true for the root and thematic formations. Also, looking at later Middle Indic shows that the nonsigmatic aorists have disappeared altogether. Therefore, there is a clear competition between the sigmatic and nonsigmatic formations, a competition which the sigmatic eventually wins. Since the two types of formations are semantically identical, they appear in (mostly) the same contexts, and thus a simple two–way contrast can be counted: does a past–tense sentence attest a sigmatic or a nonsigmatic aorist? For any given text the proportion of sigmatic to nonsigmatic aorists can be determined and this proportion used as the basis of the induced chronology.

III. Methodology.

As mentioned before, the Pali Canon comprises some four million words. There are a number of traditional divisions of these texts, either by style into nine types or into the

Tipitaka ("Three Baskets") and their subdivisions, reflecting a more semantic or functional division. The latter is more commonly used. The Pali Canon was transmitted orally for several centuries and experienced considerable modification during that time, as divisions and even individual texts were expanded and abbreviated. Therefore, this study has taken the traditional divisions and further refined them into units which have a reasonable chance of stemming from a single historical period. This results in somewhat over 10 000 texts to be dealt with, ranging in size from many two–liners³ to one text of over 20 000 words.⁴ Of these texts, slightly over 5000 attest at least one aorist of any formation, providing a good source of data for this study.

Unfortunately, most of these texts cannot be used because they do not attest aorists robustly enough. That is, while the proportion between (eg) 15 sigmatic and 5 nonsigmatic aorists is significant, the difference between one sigmatic and no nonsigmatics is not trustworthy. For example, consider a text consisting solely of two lines of verse, attesting only a single sigmatic aorist. Was that form used because it was the only formation available at the time the poem was written, or because it fit the metre while the corresponding nonsigmatic did not? Such a short text does not allow that question to be answered with any confidence. Therefore, any text which attested fewer than 10 aorists was discarded from the study.⁵ This left 685 texts, attesting a total of 27145 tokens or an average of 40 aorists per text.⁶ For each text the number of sigmatic and nonsigmatic aorists was counted, and the proportion of sigmatic aorists to total aorists was calculated. The average proportion was somewhat over 0.76, indicating that about three out of every four aorists in the set of texts were sigmatic. This is perfectly consistent with the idea that the sigmatic aorists are the productive and winning formation in Pali.

A working assumption was made that this calculated proportion was a perfect correlate to historical age, and that any text which showed a low proportion was necessarily earlier than any text with a higher proportion. This is at best only approximately true but allows the rest of the analysis to proceed. The texts are ordered by proportion and plotted as Figure 2.

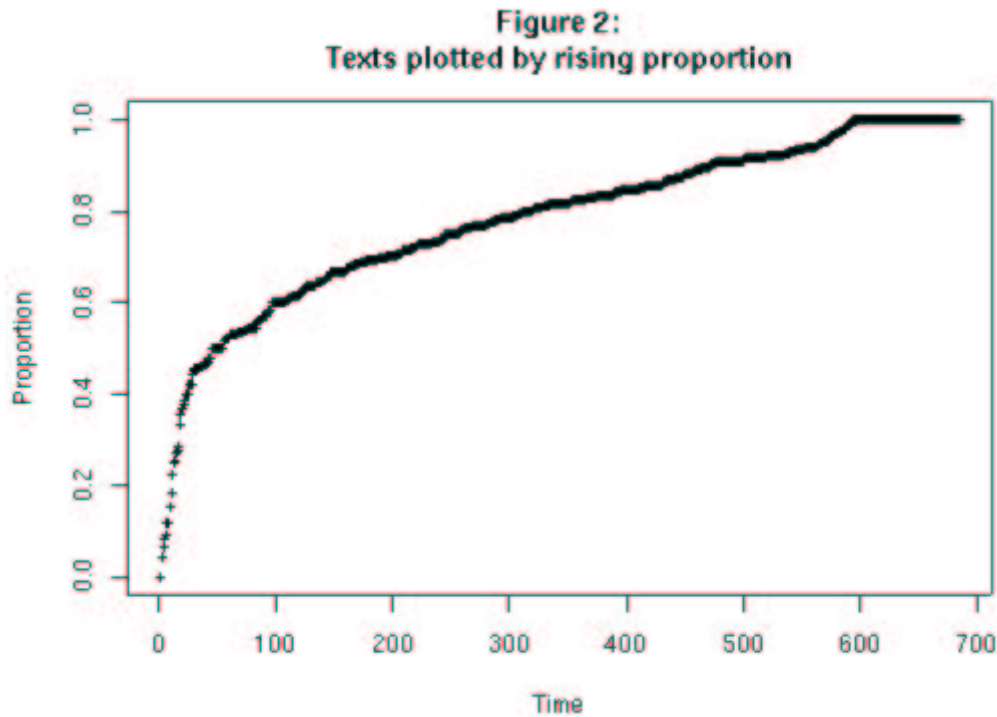
The first thing that is remarkable about this plot is that it does not look much like the expected sigmoid. Specifically, the beginning of the curve rises quickly, rather than the middle as would be expected. This indicates that the time dimension as plotted is inaccurate, and some portions need to be stretched and others compressed. In terms of the chronology of the texts, the assumption had been made with this plot that texts were

3 Many of the metrical texts are poems of only a single verse, and the traditional scholarship recognizes that many of these poems were written by distinct authors.

4 The Yamaka, an exhaustive treatise on dualism.

5 Why 10? Why not?

6 This number is slightly skewed by two texts which attested over 1000 aorists each, but since the average number of aorists per text is unimportant to the study as a whole, this is acceptable. Removing these two texts from the calculation still leaves an average of almost 31 aorists per text.

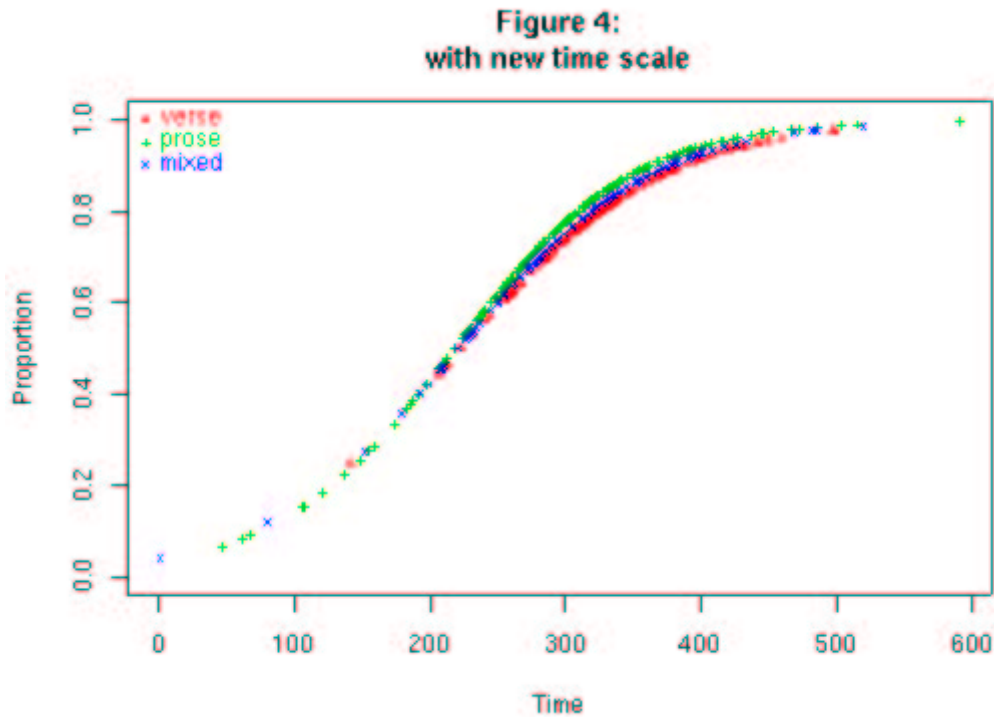


written at a constant rate (say, one per year). This is clearly false. It is therefore necessary to adjust the assumed date of composition to better fit the expect curve.

Before this is done, however, it is useful to consider the effect of context. Part of the Constant Rate Effect is the assumption that the context of a particular grammatic feature governs the rate of change. A useful division of contexts is the difference between metrical text (verse) and straight prose. Both these types are well attested in Pali and in the selected texts of this study. Somewhat over half of the texts used are just prose and about 13% are verse. The remaining third are texts where the verse and the prose is intermixed. It is expected that these three types will show different curves, reflecting different usage of the innovative sigmatic aorists in the different contexts.

In order to determine an improved chronology, the constants from the logistic equation must be determined. This is done relatively easily by assuming the plotted curve is a "noisy" or partially inaccurate reflection of the true curve, which can be found with a logistic regression, a standard statistical function⁷. Because of the expected difference between verse and prose, three separate logistic regressions are run, resulting in the three curves seen in Figure 3.

⁷ My thanks to Sasha Popescul for his help with this part of the statistics.

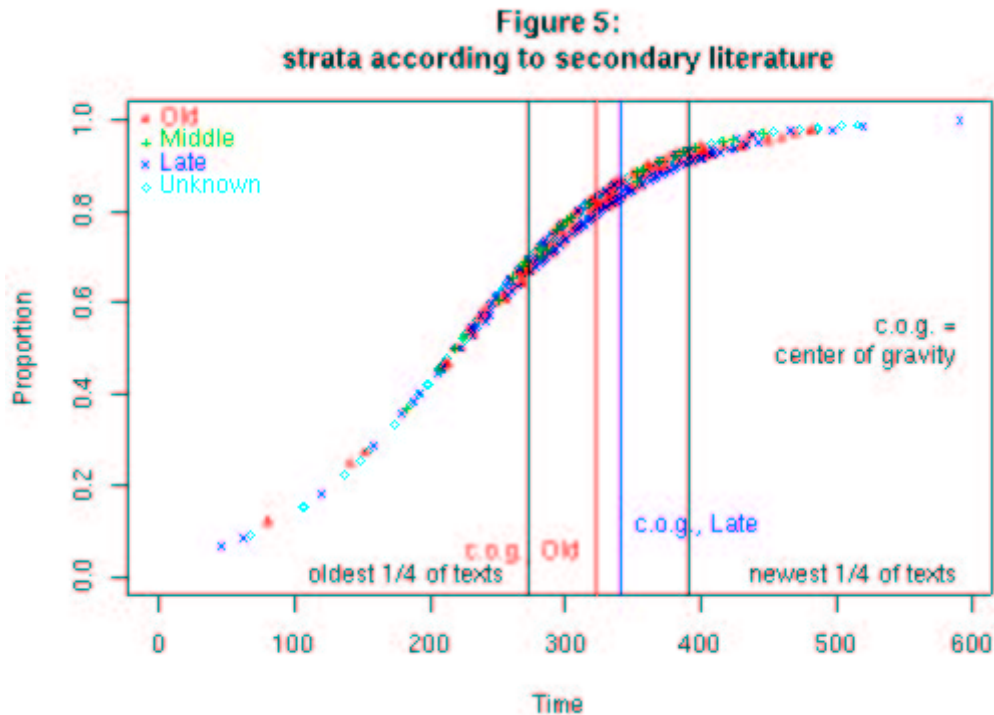


is also undefinable, sometimes taken to be negative infinity. Thus, the ends of the graph are pushed out to infinite values. While this might seem to be a mistake or a byproduct of the statistics, it actually reflects the linguistic reality quite well. There is always a small chance that a nonsigmatic aorist could be used, even at a very late date, indicating that the end point of the grammatical change never actually arrives.⁸ The inverse case is harder to understand intuitively; it suggests that there is always a possibility of producing a sigmatic aorist even before the innovation begins.⁹ Fortunately there are not many of these cases—only 3, one from each of the text types. The remaining (sensical) time numbers are scaled and recentered to conform to the old time numbers, and plotted into Figure 4.

We now have the classic sigmoid curve. Note, however, how the graph is densest somewhat to the right and above the halfway point—timestamps 200 to 400, approximately, or with proportion values between 0.6 and 0.9. This is to be expected. The movement towards the sigmatic aorist had been ongoing by the time of the Pali Canon, having begun even back in Old Indic. What is happening in Pali is the logical

8 In terms of a learning paradigm of linguistic change, there is always a chance that a child learner will hear a nonsigmatic aorist, incorporate it into her own language, and pass it on to the next generation. This chance becomes vanishingly small, but never completely disappears.

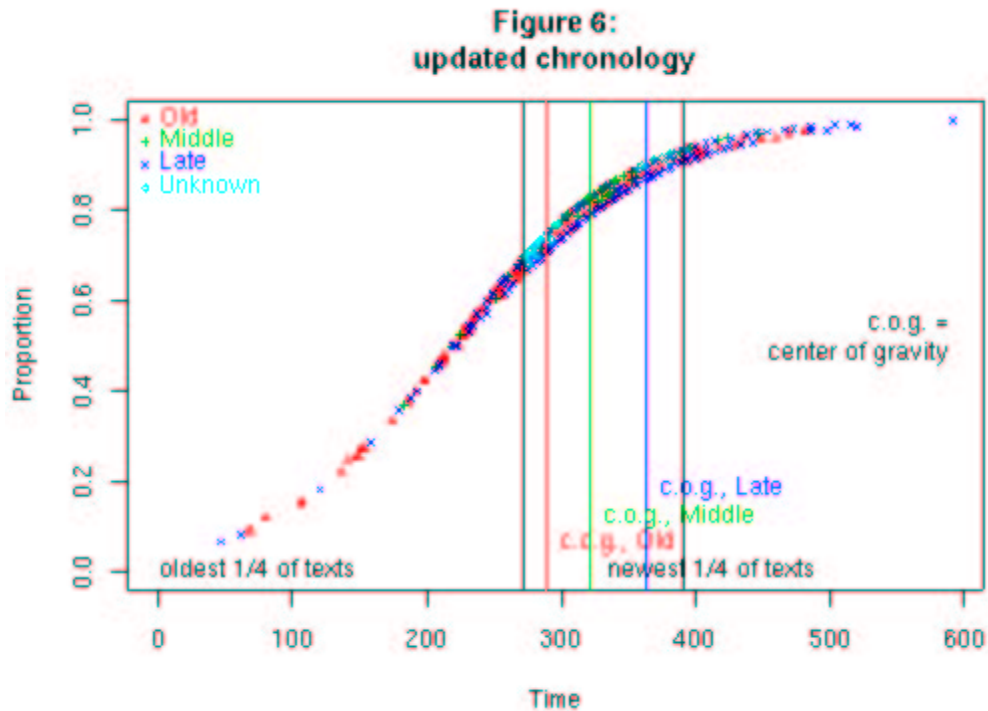
9 This means effectively that it is (slightly) possible to create a sigmatic aorist, even before that becomes a grammatic formation in the language!



conclusion of this change. The differences between the three types of text are emphasized as well.

The open question at this point is how this relates to other received wisdom about the chronology of the texts under consideration. There has been a fair amount of research (eg, Winternitz 1956, Pande 1974, Law 1974, Warder 1980, Norman 1983, von Hinüber 1996) concerning the development of the Pali Canon, although these works are hindered either by their lack of methodological rigor or their narrow scope. These scholars have offered chronologies based on other aspects of Pali linguistics, textual criticism, theology, and internal evidence such as quotations and citations within the Canon. The coordination of these criteria is no small matter. Nevertheless, they do offer a guide to the possible chronology of the texts. On the basis of these studies I have partitioned the Canon into three strata, "old" "middle" and "late", with an additional layer of "mixed" for those texts which are probable combinations of material from multiple strata. Details on this stratification can be found in Kingsbury (2002). Naturally, there is a sizable residue of texts which cannot be assigned to any stratum and are labelled as "unknown." The texts under consideration in this study fall into the five strata in roughly the expected proportions. It is the "old" "late" and "unknown" texts which are the most interesting.

As seen in Figure 5, the "unknown" texts dominate the field, comprising 206 of the 695 texts under consideration. The vertical lines indicate the "centers of gravity" for all texts and the texts classified as "old" or "late". Note that the "late" texts are appearing, on



average, slightly later than the "old" texts, seen by the fact that the median of the "old" texts is to the left of the median of the "late" texts.¹⁰ This difference can be exploited for reclassifying the "unknown" texts. The idea is simple: once again, the assumption is made that the proportion of sigmatic aorists is an indicator of age. The "unknown" texts with a clearly low proportion, then, can be reclassified as belonging to the "old" stratum and similarly with high proportions and the "late" stratum. The only problem is the definition of "clearly," and that is easily solved by selecting only the outlying quarters of texts as defined by the induced timestamps. These divisions are indicated in Figure 5 by the black vertical lines. It might seem more intuitive to use thirds rather than quarters, since we are trying to divide the texts into three strata rather than four. This would be the wrong approach, however, since it is not clear that texts placed near the (eg) one-third mark should necessarily belong to one stratum or the other. There is a margin of error, in other words, and using the one-third and two-thirds divisions would fall within that margin. Dividing by quarters, on the other hand, avoids this margin of error and the hypothesized stratification is more trustworthy. The same applies for texts in the middle—only the middlemost quarter may be trusted. With these caveats in place, however, most of the "unknown" texts can be reclassified, resulting in the distribution seen in Figure 6.

¹⁰ The "middle" texts fall in place between the "old" and "late", but for the sake of clarity that line is omitted.

This process allows 141 of the 180 "unknown" texts to be assigned to a stratum. These texts are:

<i>Old</i>	<i>Middle</i>	<i>Late</i>
Sn 2.7 brāhmaṇadhammika	DN–5.16	Vin.Cullavagga 2
DN 20.10	DN 16.i.18	Vin.Cullavagga 3
MN 2 sabbāsava	DN 16.i.23	DN 5.23
MN 14 cūḷa– dukkhakkhandha	DN 16.ii.31	DN 5.25
MN 23 vammika	DN 19.13	MN 4 bhayabherava
MN 27 cūḷa– hatthipadopama	MN 1 mūlapariyāya	MN 25 nivāpa
MN 31 cūḷagosiṅga	MN 13 mahā– dukkhakkhandha	MN 30 cūḷasāropama
MN 32 mahāgosiṅga	MN 34 cūḷagopālaka	MN 68 naḷakapāna
MN 54 potaliya	MN 53 sekha	MN 151 piṇḍapātapāri
MN 56 upāli	MN 60 apaṇṇaka	AN 4.2.2.8.
MN 59 bahavedaniya	MN 72 aggi–vacchagotta	AN 4.5.5.2.
MN 79 cūḷasakuludāyi	MN 73 mahā–vacchagotta	AN 5.2.1.5
MN 87 piyajātika	MN 76 sandaka	AN 5.4.5.6
MN 94 ghoṭamukha	MN 83 makhādeva	AN 6.1.2.7.
MN 99 subha	MN 89 dhammacetiya	AN 7.1.4.11
MN 101 devadaha	MN 95 caṅkī	AN 7.1.4.12
MN 104 sāmagāma	MN 97 dhānañjāni	AN 7.2.2.9
MN 126 bhūmija	MN 100 saṅgārava	AN 9.1.2.2
MN 140 dhātuvibhaṅga	MN 128 upakkilesa	AN 9.1.2.10
MN 145 puṇṇovāda	MN 146 nandakovāda	AN 9.1.4.8
AN 2.2.5.	MN 147 cūḷarāhulovāda	AN 9.1.4.10
AN 2.4.5.	AN 3.1.2.5.	AN 10.1.1.10
AN 3.1.4.5.	AN 6.1.2.6.	SN 1.2.30
AN 3.2.1.10.	AN 6.1.5.2.	SN 2.3.17–32
AN 3.2.2.4.	AN 6.1.5.12.	SN 5.3.4.1
AN 3.2.4.11.	AN 6.2.1.1.	SN 5.7.1.9
AN 3.3.3.4.	AN 7.1.5.9	SN 5.7.2.7
AN 3.3.3.6.	AN 7.2.2.2.	SN 5.12.2.11–12
AN 4.1.3.1.	AN 8.1.2.2.	
AN 4.4.1.9.	AN 8.1.3.10	
AN 5.3.5.3	AN 9.1.2.9	
AN 5.4.2.6	AN 10.2.3.5	
AN 6.1.2.3.	AN 10.2.5.4	
AN 6.1.5.7.	AN 10.2.5.6	
AN 6.2.1.2.	AN 10.3.2.3	
AN 6.2.1.8.	AN 10.4.2.6	

<i>Old</i>	<i>Middle</i>	<i>Late</i>
AN 7.2.1.3.	SN 1.10.8	
AN 7.2.1.9	SN 2.5.10	
AN 8.1.2.10.	SN 2.10.3	
AN 8.1.6.1	SN 5.2.6.2	
AN 8.1.8.3	SN 5.2.6.3	
AN 9.1.1.3	SN 5.2.6.4	
AN 9.1.2.1	SN 5.3.1.9	
AN 10.1.3.6	SN 5.3.1.10	
AN 10.2.4.5	SN 5.11.3.6	
SN 1.1.20		
SN 1.10.12		
SN 2.5.6		
SN 2.5.11		
SN 2.8.1		
SN 2.8.21		
SN 2.10.4		
SN 2.10.10		

It is important to remember that this is only a tentative chronology for these texts. The single criterion of the aorists, while more compelling than the more subjective tests, cannot be taken as the last word and must be combined with other judgements. The texts above are thus presented as a list of good candidates for further examination.

IV. Accuracy

In order to make any estimates about the accuracy of this methodology a little sleight-of-hand needs to take place. That is, in an ideal world the proper stratification for each of the "unknown" texts would be known and thus that stratification could be compared to the one reached by this method. But of course, it is the very fact that the texts are of unknown date that necessitates this method in the first place! Instead, a variant of 'held-out estimation' can be used. In this strategy, a number of texts which have been assigned to some stratum are temporarily reassigned to the "unknown" stratum and the procedure run as before. These 'held-out' texts will then be assigned to a stratum just like any other "unknown" stratum text, and this assignment can be compared to the 'correct' stratification previously given. Repeating this several times, with different texts held out for testing, allows for a good estimation of the accuracy of the procedure on the truly unknown texts.

Results, while hardly overwhelming, are encouraging. On average, texts are assigned to the same stratum on the basis of this methodology and on the more subjective criteria 24% of the time. However, two of the strata used by the subjective method are "mixed" and "unknown", which this method does not even make an attempt at. It is therefore incorrect to grade this method on those texts. Removing those texts from the test raises

the accuracy to about 58%. Furthermore, errors tend to be relatively minor in scope—that is, a text which was assigned to the "old" stratum on the basis of the subjective criteria might be assigned to the "middle" stratum based on this methodology, but it will usually not be assigned to the "late" stratum.

It must be remembered, also, that the target stratification is not without error itself. Many of the texts which were supposedly misclassified by this methodology have chronologies which are suspicious at best. For example, the texts in the Niddesas are all assigned to the "old" stratum on the basis of their being commentaries on the Suttanipata, an undeniably old text. However, being a commentary on an old text does not require that the Niddesas themselves be old, and indeed many of the Niddesas were restratified into the "middle" or "late" strata. Similarly, the texts in the Suttavibhanga, being combinations of old rules and later explanations, are a difficult target and should realistically be assigned to a "mixed" category rather than "old." On the converse, texts which clearly belong to one stratum or another are usually correctly assigned to that stratum. For example, the texts of the Suttanipata are almost always assigned to the "old" stratum, rarely the "middle" stratum, and never to the "late," while the Apadanas are usually assigned to the "late" stratum, with the remainder falling in the "middle" stratum.

V. Conclusion.

A little simple mathematics and a single assumption allows for a large number of previously unclassified texts to be assigned to a chronological stratum, by taking a small difference and exaggerating it. In theory, it is possible to use the same methodology to change all of the stratification, but it must be remembered that this study used only one grammatical change out of many, and that the chronology used was based only a probabilistic distribution. That is, it is probable that a text with a low proportion of sigmatic aorists comes from an earlier stratum, but is not necessary for that to be the case. In order for an already-stratified text to be reclassified, effectively overriding the knowledge of experts of the field, a large number of (unrelated) changes would have to be compared and the chronologies proposed by each of those changes combined. With the methodology of this paper in place, however, that is a relatively simple operation and the resulting chronology of the Pali Canon should be both wide in scope and methodologically rigorous.

V. References

- Bloch, Jules. 1877/1965. Indo–Aryan: From the Vedas to Modern Times. English edition, trans. Alfred Master. Paris: Librairie d’Amerique et d’Orient.
- Fahs, Achim. 1985. Grammatik des Pali. Leipzig: VEB Verlag Enzyklopädie.
- Geiger, Wilhelm. 1956. Pali Literature and Language. Second Edition, trans Batakrishna Ghosh. Calcutta: University of Calcutta Press.
- Hinüber, Oskar von. 1977. Notes on the e–preterite in Middle Indo–Aryan. *Münchener Studien zur Sprachwissenschaft* 36: 39–48.
- Hinüber, Oskar von. 1994. Selected Papers on Pali Studies. Oxford: Pali Text Society.
- Hinüber, Oskar von. 1996. A Handbook of Pali Literature. Berlin: Walter de Gruyter & Co.
- Kingsbury, Paul. 2002. The Chronology of the Pali Canon: The case of the aorists. PhD dissertation, Department of Linguistics, University of Pennsylvania.
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199–244.
- Law, B.C. 1974. History of Pali Literature. Varanasi: Bhartiya Publishing House.
- Norman, K.R. 1983. Pāli Literature. Wiesbaden: Otto Harrassowitz.
- Pande, Govind Chandra. 1974. Studies in the Origins of Buddhism. Second Edition. Delhi: Motilal Banarsidass.
- Santorini, Beatrice. 1993. Phrase structure change in Yiddish. *Language Variation and Change* 5: 257–283.
- Taylor, Ann. 1994. The change from SOV to SVO in Ancient Greek. *Language Variation and Change* 6: 1–37.
- Warder, A.K. 1980. Indian Buddhism. Second edition. Delhi: Motilal Banarsidass.
- Winternitz. 1972. A History of Indian Literature. Vol II, second edition. New Delhi: Oriental Books Reprint Corporation.