

# Dating Tahitian Oral Tradition Texts from Primarily Internal Evidence

David Meyer  
University of Edinburgh

david.meyer@ed.ac.uk

Over 220 Tahitian oral tradition texts, representing about 75,000 words, made their way into print during the 19th and early 20th centuries. Most of the material seems to have originated before contact with Europeans, although a small amount may date to after the introduction of Christianity.

Some of the texts appear to reflect several oral poetic styles of the pre-Contact and pre-literate era. Others are more prose-like. Yet others lie in between, with somewhat modern-sounding language occasionally punctuated by poetic structure. In order to obtain insight into the language, culture, and oral poetry of pre-Contact Tahiti, it would be useful to limit study to earlier texts, thus avoiding material that may represent re-analysis, external influence, or post-Contact stylistic evolution.

Recitation dates are available for about 30% of the texts. A method will be proposed here for determining rough dating for the other 70%. The method will be developed from two types of internal analysis. Once trained on the dated texts, the same method should be applicable to undated texts, as there is nothing in the source material to suggest that date omission was anything other than arbitrary.

## 1. Introduction

Over 220 Tahitian oral tradition texts, representing about 75,000 words, made their way into print during the 19th and early 20th centuries. Most of the material seems to have originated before contact with Europeans, although a small amount may date to after the introduction of Christianity.

Some of the texts appear to reflect several oral poetic styles of the pre-Contact and pre-literate era. Others are more prose-like. Yet others lie in between, with somewhat modern-sounding language occasionally punctuated by poetic structure. In order to obtain insight into the language, culture, and oral poetry of pre-Contact Tahiti, it would be useful to limit study to earlier texts, thus avoiding material that may represent re-analysis, external influence, or post-Contact stylistic evolution.

Recitation dates are available for about 30% of the texts. A method will be proposed here for determining rough dating for the other 70%. The method will be developed from two types of internal analysis.

The first analysis will consist of an exhaustive search through the dated texts to identify good examples of early oral poetry. Half of the material thus identified will be compiled into an early oral poetic corpus.

In a subsequent analysis, all of the other dated oral texts will be compared to the early oral corpus and, for maximal contrast, to a corpus of modern prose. Participating in the comparison will also be nine modern Tahitian poems, and five texts of early 19<sup>th</sup> century prose.

Finally, a method of rough dating will be devised based on analysis of the similarity between individual dated texts and the two corpora.

Oral tradition was closely linked to the practice of Tahitian religion, which was formally abolished in 1815. Following this, the relevance of pre-literate and pre-Christian oral tradition would presumably have been on the wane. For the purposes of separating the oral poetry of the professional TAHU'A class from what would follow, a somewhat arbitrary year of 1850 is proposed.

It is likely that some of the texts recited after 1850 are good examples of an early style. However contextual information for them is sparse, and there would seem to be mounting risk that an oral poet reciting after 1850 would never have trained or performed in an early religious capacity.

## 2. Analysis

### 2.1 Analyzing the early texts for poetic structure

It will be assumed that early texts that exhibit a density of poetic structure are safer to use as examples of early poetry than less structured texts; although the latter may still be representative of an early style.

Many of the pre-1850 texts are densely poetic. Poetic patterns include, but are not limited to, repetition of surface form, repetition of a combination of surface form and syllable count or of surface form and part-of-speech category, repetition of phonemes and/or phonemic features, semantic repetition, and often chiasmatic variations of the above.

If a text has several good examples of just one of these types of pattern, then it will be considered poetic. If the same text includes examples of multiple types, then it will be considered to be very poetic.

There are 58 dated<sup>1</sup> oral texts that consist of at least 50 words<sup>2</sup>, and that are not near duplicates of others texts. All of these were exhaustively analyzed for good examples of two specific types of poetic pattern: repetition of surface form and syllable count, and repetition of surface form and part-of-speech category. These two pattern types were chosen because they lent themselves well to computationally-assisted detection. A description for each type is provided in sections 2.2 and 2.3.

A text is not necessarily unpoetic because it contains neither of these particular types of pattern. On the other hand, if a text were to contain good examples of both, it should be suitable for inclusion into the corpus. It would be unlikely for a pre-1850 text to be poetically structured in more than one way, but not be representative of early oral poetry.

### 2.2 Examples of the surface form and syllable count pattern

This type of pattern exhibits repetition of surface form, syllable count, and often syntactic pause. The overall syllable count will either remain the same, increment, or decrement from one member of the pattern to the next.

In (1)<sup>3</sup> we encounter an example where syllable count remains the same:

---

<sup>1</sup> If more than one recitation date exists for a text, an average date is calculated.

<sup>2</sup> Analysis results are inconsistent for very small texts, and especially for those containing fewer than 50 words.

<sup>3</sup> In this and the other examples, the first line will contain the IPA form, as well as an indication of syntactic pause. The second line will provide either syllable count or part-of-speech category information, depending on the type of poetic pattern described. The third line is a morpheme-by-morpheme gloss. The fourth line is the translation provided in the published text.

To assist in identifying the poetic pattern, repetition of surface form and pause are indicated in bold, and words corresponding to syllable count or part-of-speech category repetition are displayed in italics.

The grammatical abbreviations are: 3S: 3rd person singular, ADV: adverb, ART: article, CONT: continuative aspect, DIROBJ: direct object marker, EXIST: existential particle, IMPERF: imperfect aspect, MODIF: noun or adjective acting as a modifier, NEUT.ALIEN: neutral alienable possession, PERF: perfect aspect.

(1) Extract from ‘Warning by messengers of the pa’i-atua service’ (Henry 1928:158-159):

1. PAUSE e            noho    i            niʔa    i            te    ma:hora  
           1            2            1            2        1            1        3  
           IMPERF    sit            DIROBJ    on        DIROBJ    the     yard  
           And sit upon the lawn
2. PAUSE e            noho    e            ʔupu            i            te    ʔaho:ʔa  
           1            2            1            2                    1            1        3  
           IMPERF    sit            IMPERF    recite-prayer    DIROBJ    the     brush-clearing-prayer  
           To recite the *ahoa* (life giving)

The pattern that repeats here is:

PAUSE e noho 8 syllables

The example in (2) exhibits decrementing syllable count:

(2) Extract from ‘Te parau a Honoura’ (Henry 1895:256-291):

1. PAUSE tiʔa            aʔe        ra        te        ʔie  
           2            2            1        1        2  
           stand        upward    there    the     sail  
           The sails were set,
2. PAUSE mau            aʔe        ra        te        hoe  
           1            2            1        1        2  
           fastened    then        there    the     paddle  
           the paddle guided,
3. PAUSE te:    tiʔa        ra        te        ʔie  
           1        2            1        1        2  
           CONT stand    there    the     sail  
           The sails were still set,
4. PAUSE mau            aʔe        ra            te        hoe  
           1            2            1            1        2  
           fastened    upward    there        the     paddle  
           and the paddle guided,
5. PAUSE te:        tiʔa        ra            te        ʔie  
           1            2            1            1        2  
           CONT stand    there        the     sail  
           The sails were still set,

6. PAUSE *te: mau ra te hoe*  
 1 1 1 1 2  
 CONT fasten there the paddle  
 the paddle still guided,

The pattern that is repeating here is:

PAUSE *variable syllable count te ?ie* PAUSE *variable syllable count te hoe*

where the combined variable syllable count for each two line pattern member decrements from 9 to 8 to 7.

### 2.3 Examples of the surface form and part-of-speech category pattern

This type of pattern exhibits repetition of surface form, part-of-speech category, and often syntactic pause. Below are two examples:

(3) Extract from ‘The Chaotic Period’ (Henry 1928: 340-344):

1. PAUSE **e one i te** *ʔaxere nuʔu*  
 EXIST NOUN DIROBJ ART NOUN MODIF  
 EXIST sand DIROBJ the space army  
 sand in the space for armies,
2. PAUSE **e one i te** *fenua mahora*  
 EXIST NOUN DIROBJ ART NOUN MODIF  
 EXIST sand DIROBJ the land open  
 sand on the plains,
3. PAUSE **e one i te** *ʔa.na: vai*  
 EXIST NOUN DIROBJ ART NOUN MODIF  
 EXIST sand DIROBJ THE riverbed freshwater  
 sand in the river beds,
4. PAUSE **e one i te** *tororaʔa mouʔa*  
 EXIST NOUN DIROBJ ART NOUN MODIF  
 EXIST sand DIROBJ the stretching mountain  
 sand for the mountain ranges,
5. PAUSE **e one i te** *ʔaxere nuʔu ra:ʔau*  
 EXIST NOUN DIROBJ ART NOUN MODIF MODIF  
 EXIST sand DIROBJ the space army tree  
 sand for the forest wilds.

The fairly simple pattern repeated here is:

**PAUSE e one i te** NOUN MODIF

(4) Extract from ‘The deluge, Tahitian version’ (Henry 1928: 445-448)<sup>4</sup>:

1. **PAUSE ?ua rave a?e ra te vahine**  
 PERF VERB ADV ADV ART NOUNHUMAN  
 PERF take then there the woman  
 The woman took
2. **i ta:-na fanau?a ri?i moa**  
 DIROBJ NEUT.ALIEN-3S NOUN.SMALLBEING MODIF MODIF.ANIMAL  
 DIROBJ NEUT.ALIEN-3S newborn small chicken  
 her little chickens,
3. **PAUSE ?ua rave iho ra te tane**  
 PERF VERB ADV ADV ART NOUN.HUMAN  
 PERF take just there the man  
 the man took
4. **i ta:-na pinia ri?i pua?a**  
 DIROBJ NEUT.ALIEN-3S NOUN.SMALLBEING MODIF MODIF.ANIMAL  
 DIROBJ NEUT.ALIEN-3S small.animal small pig  
 his young pigs;
5. **PAUSE ?ua rave atu ra te vahine**  
 PERF VERB ADV ADV ART NOUN.HUMAN  
 PERF take thither there the woman  
 the woman took
6. **i ta:-na fanau?a ri?i uri**  
 DIROBJ NEUT.ALIEN-3S NOUN.SMALLBEING MODIF MODIF.ANIMAL  
 DIROBJ NEUT.ALIEN-3S newborn small dog  
 her young dogs

The more complex pattern found in each two line member of (4) is:

**PAUSE ?ua rave** ADV **ra te** NOUN.HUMAN **i ta:-na** NOUN.SMALLBEING **ri?i** MODIF.ANIMAL

<sup>4</sup> Note that the apparent semantic restrictions have been included in part-of-speech category labeling.

## 2.4 Contents of the early oral poetic corpus

In the 58 oral texts analyzed, instances of surface form and part-of-speech category repetition turned out to be more common than repetition of surface form and syllable count. In all, 19 texts were found to contain good examples of both. Of these, 17 pre-dated 1850.

Eight of the 17 texts, representing 9,667 words, were selected at random for inclusion into the early oral poetic corpus<sup>5</sup>. The other nine were set aside for testing. A procedural flaw would be implied if the texts reserved for testing were not found to be similar to those constituting the corpus.

## 2.5 Contents of the modern prose corpus

As much as the early oral poetic corpus is early and poetic, it was intended that its counterpart be prose-like and modern. A 10,379 word corpus of modern prose<sup>6</sup> was compiled containing material from the following types of sources:

Internet blog sites: 4,115 words  
 Website articles: 3,457 words  
 Oral history: 791 words  
 Student essays: 2,016 words

## 2.6 Lexical vector space analysis

The next task will be to compare against these two corpora the nine oral poetic texts reserved for testing, all of the other oral tradition texts, the five early 19th century prose texts, and the nine modern poems. The comparison will attempt to determine to which corpus each text is most similar, and to what degree. Lexical vector space analysis will be the means employed to accomplish this<sup>7</sup>.

In this type of analysis, each word type<sup>8</sup> in a text corresponds to a dimension of vector space, where the word's number of tokens is reflected by that vector's length.

A multidimensional vector is calculated that represents all instances of all words in the text. One text's multidimensional vector is compared to that of another to determine document similarity. If both texts are identical, then the angle between their two vectors will be 0°.

The angle between vectors is often represented by its cosine value. As the cosine of 0° is 1, another way of gauging similarity is by how close the cosine is to 1.

---

<sup>5</sup> The contents of the early oral poetic corpus are Ahnne (1924:20-23), Henry (1928:306-307), Henry (1928:336-338), Henry (1928:340-344), Henry (1928:364-371), Henry (1928:409-413), Henry (1928:413-415), and Henry (1928:445-448).

<sup>6</sup> The contents of the modern prose corpus are Aorai (2004), Arahau (2004), Porinetia ... (2003), Te ma'i aho pau (2003), Te mau parau 'āpī ... (2006), Te oraraa taatiraa (n.d.), Teri'iama (2001), TKNui (2003a), TKNui (2003b), and Yon Yuc (1997).

<sup>7</sup> In synchronic vector space comparisons, a stopword list of function words is commonly subtracted from each document to be compared. This has the effect of both speeding up computational processing and permitting content words to take the lead in establishing similarity. A stopword list will not be applied in the current analysis, however, because how function words are used is likely to carry increased significance for texts that span time period and genre.

<sup>8</sup> In the current analysis, the surface form of a word will serve as the vector space term. In a different test, a term could just as easily represent metadata (e.g. a word's part-of-speech category), or information about more than one word (e.g. a bigram or trigram of adjoining words).

## 2.7 Results of comparing the dated texts to both corpora

The vector space similarity results are shown in table 1. The texts have been sorted first by the corpus to which they are most similar, and next by the absolute value of the difference between the cosines. Difference between cosines measures the contrast between a text's similarity to each corpus.

Note that the sort order switches from greatest to least contrast with the first corpus to least to greatest with the second. This is so that results may be viewed as a continuum, with texts that are most like modern prose at the top, and those most like early oral poetry at the bottom. At the boundary of corpus similarity (between lines 17 and 18) can be found texts that are not strongly similar to either of the corpora.

**Table 1. Vector space comparison of the dated texts to both corpora**

Texts that are more similar to the modern prose corpus (from greatest to least difference between cosines):

	<u>Difference</u> <u>between</u> <u>cosines</u>	<u>Cosine to</u> <u>modern prose</u> <u>corpus</u>	<u>Cosine to</u> <u>early oral</u> <u>poetic corpus</u>	<u>Text</u>	<u>Genre</u> <sup>9</sup>	<u>Word</u> <u>count</u>	<u>Date</u>
1.	0.09013126	0.55940172	0.46927046	Raapoto (1990)	Modern poetry	84	1990
2.	0.08366750	0.70190035	0.61823284	Pomare II (1817b)	19th century prose	437	1817-07-03
3.	0.06938043	0.63275333	0.56337290	Pomare II (1812)	19th century prose	474	1812-10-12
4.	0.06931203	0.79479038	0.72547834	Burau and Miro (1836)	19th century prose	5953	1836
5.	0.06660333	0.68238037	0.61577704	Pomare II (1817a)	19th century prose	470	1817-07-03
6.	0.06075113	0.67930642	0.61855529	Mapuhi (1985:7)	Modern poetry	399	1985
7.	0.04285871	0.67079731	0.62793859	Mapuhi (1985)	Modern poetry	298	1985
8.	0.03372284	0.73599105	0.70226821	Pomare II (1825)	19th century prose	10811	1825
9.	0.03156323	0.64871319	0.61714995	Mapuhi (1993)	Modern poetry	248	1993
10.	0.02582726	0.62560330	0.59977603	Henry (1928:244)	Other oral tradition	244	1843
11.	0.02540712	0.64668871	0.62128159	Caillot (1914:131-141)	Other oral tradition	2113	1912-1913
12.	0.01746490	0.61677080	0.59930589	Henry (1928:245-246)	Syllable and POS	309	1843
13.	0.01611982	0.55916690	0.54304707	Raapoto (2005)	Modern poetry	84	1990
14.	0.01118132	0.59636450	0.58518317	Caillot (1914:117-124)	Other oral tradition	1564	1912-1913
15.	0.00720662	0.55465174	0.54744511	Henry (1928:522-523)	Other oral tradition	119	1896
16.	0.00252378	0.53529691	0.53277312	Caillot (1914:111-113)	Syllable and POS	334	1912-1913
17.	0.00243395	0.53722687	0.53479292	Henry (1928:521)	Other oral tradition	140	1896

Texts that are more similar to the early oral poetic corpus (from least to greatest difference between cosines):

	<u>Difference</u> <u>between</u> <u>cosines</u>	<u>Cosine to</u> <u>modern prose</u> <u>corpus</u>	<u>Cosine to</u> <u>early oral</u> <u>poetic corpus</u>	<u>Text</u>	<u>Genre</u>	<u>Word</u> <u>count</u>	<u>Date</u>
18.	0.00461910	0.63296216	0.63758127	Hiro (1991)	Modern poetry	562	1979
19.	0.00599098	0.52134528	0.52733627	Alexander (1893:57)	Other oral tradition	101	1881
20.	0.00639702	0.58355562	0.58995264	Hiro (n.d.:57)	Modern poetry	312	<= 1990
21.	0.00758881	0.50014576	0.50773458	Caillot (1914:114-116)	Other oral tradition	391	1912-1913
22.	0.00849566	0.54143946	0.54993512	Henry (1928:334)	Other oral tradition	239	1891

<sup>9</sup> Texts that were found to contain both types of poetic pattern described above are labeled 'Syllable and POS'. All of the other oral tradition texts are simply labeled 'Other oral tradition'. The five early 19<sup>th</sup> century prose texts are labeled '19<sup>th</sup> century prose', and the nine examples of modern Tahitian poetry 'Modern poetry'.



23.	0.00953645	0.46134495	0.47088141	Brémond (1982)	Modern poetry	97	1982
24.	0.01553497	0.49077715	0.50631213	Henry (1897:211-212)	Other oral tradition	165	1897-08
25.	0.01705130	0.47293756	0.48998886	Alexander (1893:59)	Other oral tradition	79	1881
26.	0.01914107	0.60387148	0.62301256	Henry (1928:444)	Other oral tradition	251	1825
27.	0.02107677	0.60815309	0.62922986	Alexander (1893:58)	Other oral tradition	227	1881
28.	0.02179840	0.53002439	0.55182280	Devatine (2002)	Modern poetry	174	1979
29.	0.02208296	0.61438673	0.63646970	Henry (1928:332-333)	Other oral tradition	501	1818
30.	0.02478829	0.66384544	0.68863373	R.T. (1962:30-34)	Other oral tradition	1051	1961-06-09
31.	0.02513753	0.58870013	0.61383767	Henry (1928:398-399)	Other oral tradition	400	1840
32.	0.02584116	0.62503596	0.65087712	Henry (1893:106-107)	Other oral tradition	426	1890
33.	0.02673242	0.67302380	0.69975622	Henry (1928:468-470)	Other oral tradition	679	1825-12-24
34.	0.02963141	0.66119980	0.69083121	Henry (1928:429-430)	Other oral tradition	498	1825, 1885
35.	0.03010535	0.70915330	0.73925865	Henry (1928:423-426)	Other oral tradition	1131	1887
36.	0.03199984	0.44451489	0.47651473	Henry (1928:523-524)	Syllable and POS	54	1896
37.	0.03475400	0.46789913	0.50265313	Alexander (1893:58-59)	Other oral tradition	66	1881
38.	0.03508491	0.70214854	0.73723346	Henry (1928:431-433)	Other oral tradition	1000	1825, 1901
39.	0.03547747	0.53366859	0.56914607	Henry (1928:383)	Other oral tradition	117	1825, 1840
40.	0.03716379	0.52244086	0.55960465	Henry (1928:530)	Other oral tradition	132	1896
41.	0.03995933	0.49287612	0.53283546	Henry (1928:464)	Other oral tradition	128	1854
42.	0.04201064	0.68131607	0.72332671	Henry (1928:427-429)	Other oral tradition	1071	1820
43.	0.04445329	0.57894035	0.62339365	Henry (1928:461-462)	Other oral tradition	522	1886
44.	0.04761960	0.58122647	0.62884608	Henry (1928:308-309)	Syllable and POS	469	1829
45.	0.05040831	0.49730356	0.54771187	Henry (1928:531)	Other oral tradition	132	1896

Proposed precision-oriented threshold for pre-1850 dating:

	<u>Difference</u> <u>between</u> <u>cosines</u>	<u>Cosine to</u> <u>modern prose</u> <u>corpus</u>	<u>Cosine to</u> <u>early oral</u> <u>poetic corpus</u>	<u>Text</u>	<u>Genre</u>	<u>Word</u> <u>count</u>	<u>Date</u>
46.	0.05373075	0.73147744	0.78520820	Henry (1928:448-452)	Other oral tradition	1609	1822, 1824
47.	0.05966746	0.73957661	0.79924408	Henry (1895:256-291)	Syllable and POS	9265	< 1839
48.	0.06147086	0.60892952	0.67040038	Henry (1928:426)	Other oral tradition	343	1824
49.	0.06157991	0.56612553	0.62770545	Henry (1928:191-192)	Other oral tradition	485	1887
50.	0.06790852	0.51782299	0.58573151	Henry (1894:136-138)	Other oral tradition	518	1817
51.	0.06802679	0.51369593	0.58172272	Henry (1928:399-402)	Other oral tradition	632	1817
52.	0.07073401	0.67514979	0.74588381	Henry (1928:437-439)	Other oral tradition	731	1822 or 1824
53.	0.07676534	0.59329922	0.67006457	Henry (1928:395-398)	Syllable and POS	1576	1840
54.	0.08206614	0.56690860	0.64897474	Henry (1928:458)	Other oral tradition	320	1824
55.	0.08413860	0.64773796	0.73187656	Henry (1928:353-354)	Other oral tradition	719	1822, 1845
56.	0.09091719	0.66304406	0.75396126	Henry (1928:339-340)	Syllable and POS	545	1822, 1824, 1833
57.	0.09172407	0.57252477	0.66424884	Henry (1928:404-405)	Syllable and POS	315	1822
58.	0.09350753	0.56684235	0.66034989	Henry (1928:374-376)	Other oral tradition	385	1840
59.	0.09619244	0.61138372	0.70757616	Henry (1928:402-403)	Other oral tradition	533	1822, 1833
60.	0.09768603	0.52418732	0.62187336	Henry (1928:307-308)	Syllable and POS	360	1839
61.	0.10256467	0.66449471	0.76705938	Henry (1928:415-420)	Syllable and POS	1665	1833 or 1834
62.	0.10414398	0.69794574	0.80208972	Henry (1928:405-407)	Other oral tradition	1668	1825
63.	0.10954775	0.65340025	0.76294800	Emory (1938:53-58)	Syllable and POS	1269	1849-07-16
64.	0.12675481	0.61031419	0.73706901	Henry (1928:359-363)	Other oral tradition	1160	1818

We note that of the ‘Syllable and POS’ texts set aside for testing, seven of the eight that pre-date 1850, and one post-1850 text have been categorized as early oral poetic. This is an expected, and welcome, result. The two other ‘Syllable and POS’ texts were categorized as modern prose. In the case of Caillot (1914:111-113), this perhaps relates to its fairly late recitation date of 1912-1913 (see line 16). Concerning Henry (1928:245-246), however, a reason is not readily apparent (see line 12).



We find that despite their very early date, all 19<sup>th</sup> century prose texts strongly resemble the modern prose corpus (see lines 2, 3, 4, 5, and 8).

Concerning modern poetry, we note that both Raapoto poems (see lines 1 and 13) and all three Mapuhi poems (see lines 6, 7, and 9) are categorized as modern prose. However both Hiro poems (see lines 18 and 20), the Brémond poem (see line 23), and the Devatine poem (see line 28) are categorized as early oral poetic. It should be noted that Hiro and Brémond's poems are fairly close to the modern prose boundary.

One way to account for the similarity between early 19<sup>th</sup> century and modern prose is that prose may not have changed that much in the last 180 or so years. Table 1 results imply that a letter written by Pomare II prior to 1820 used language virtually indistinguishable from that of the modern-day blogger.

Another possibility is that rather than being inordinantly archaic, the language of early oral poetry employs a vocabulary that is stylistically distinct. Perhaps modern poets such as Flora Devatine are attempting to compose in an older oral poetic style whereas others, such as Rui a Mapuhi and Turo a Raapoto, are using vocabulary that is more disconnected from the early tradition.

The truth may also lie somewhere in between. Early oral poetry may consist of language that is both archaic vis-à-vis early 19<sup>th</sup> century prose, as well as being stylistically very dissimilar.

In any event, we can observe the following:

- The language of the early oral poetic corpus is quite distinct from that of early 19<sup>th</sup> century prose texts, and from the poetry of some modern poets.
- The dates for the oral tradition texts appear to form a rough timeline. From lines 1 to 45 we find that 22 of the 31 oral texts have dates of 1850 or later (70.1% precision accuracy). We can also locate 22 of the 23 post-1850 oral texts within this portion of data (95.7% recall accuracy). Moving down the list, from lines 46 through 64 we find that 18 of these 19 texts pre-date 1850 (94.7% precision accuracy). However, nine other pre-1850 oral texts were found in lines 1 through 45 (66.7% recall accuracy)<sup>10</sup>. It would appear that a precision-oriented<sup>11</sup> threshold for pre-1850 dating can therefore be established between lines 45 and 46, where the difference in cosines for line 45 happens to be .05040831, and for line 46 .05373075.

### 3. Conclusion

A proposed precision-oriented method for assigning a pre-1850 date to an undated text is as follows:

1. Using lexical vector space analysis, compare the undated text to the early oral poetic and modern prose corpora.

<sup>10</sup> Both precision and recall accuracy for pre-1850 texts are adversely affected from eight very archetypal of their number having been mustered into corpus duty. Were linguistically similar stand-ins for these eight added to the test set, precision would increase to 96.3%, and recall to 74.3%.

<sup>11</sup> A precision-oriented approach is desirable so as to minimize attribution of post-Contact developments to the pre-Contact era.

2. If the text is found to be more similar to the early oral poetic corpus, and if the absolute value of the cosine difference is greater than .052, then assign to it a date of pre-1850.

For a set of undated oral texts roughly similar to the dated texts of the training set, it is predicted that this method will achieve over 90% precision accuracy and over 65% recall accuracy. Concerning the degree of similarity between dated and undated texts, there is nothing in the source material to suggest that date omission was anything other than arbitrary.

Results from table 1 bring up an interesting possibility for further research. We noted that seven of the eight pre-1850 'Syllable and POS' texts could be found on the pre-1850 side of the dating threshold. The question might be asked of whether this is merely due to a comparison of like to like (i.e. doubly poetically structured texts compared to a corpus made up of same), or if it is an indication of a poetic density continuum that peaks with the material of the oral corpus, and then declines as early oral poetic competence fades over the course of the 19<sup>th</sup> century.

Pursuit of this question would probably first require an exhaustive analysis of the dated texts to uncover all of their poetic structure; not just the two types investigated here. If there were found to be a gradual diminishment of that structure over time, it would also be interesting to determine at what point after 1815 this first occurs, and whether the currently proposed, and very arbitrary, year of 1850 should be moved forward or back.

## References

- Ahnne, E. 1924. Le lyrisme des Tahitiens. *Bulletin de la Société des Etudes Océaniques* 9:20-23.
- Alexander, W.D. 1893. Specimens of ancient Tahitian poetry. *Journal of the Polynesian Society* 2:55-59.
- Aorai. 24 January 2004. Papatere T'ōmā/Aorai [Forum charter]. *T'ōmā Tahua 'Ōrero Mā'ohi* <<http://aorai.com/toma>>. <<http://aorai.com/toma/viewtopic.php?t=44>> [1 December 2005].
- Arahau. 2004. *Centre de Recherche et de Documentation Pédagogiques de la Polynésie Française*. <[http://www.crdp.pf/cvive\\_conc2002.htm](http://www.crdp.pf/cvive_conc2002.htm)> [1 December 2005].
- Brémond, Hubert. 1982. Tahitian poetry. *Mana* 7(1):1-86.
- Burau, and Miro. 1836. *Te hoe parau iti na Burau raua o Miro*. Tahiti: London Missionary Society Press.
- Caillot, Auguste Charles Eugène. 1914. *Mythes, légendes et traditions des Polynésiens; textes polynésiens, recueillis, publiés, traduits en français et commentés par A.-C. Eugène Caillot*. Paris: E. Leroux.
- Devatine, Flora. 1 February 2002. Fa'atara d'exhortation. *Ile-en-Ile*. <[http://www.lehman.cuny.edu/ile.en.ile/paroles/devatine\\_inedits.html](http://www.lehman.cuny.edu/ile.en.ile/paroles/devatine_inedits.html)> [1 December 2005].
- Emory, Kenneth P. 1938. The Tahitian account of creation by Mare. *Journal of the Polynesian Society* 47:45-63.
- Henry, Tuira. 1893. Te Umu-Ti, a Raiatean ceremony. *Journal of the Polynesian Society* 2:105-108
- Henry, Tuira. 1894. The birth of new lands, after the creation of Hawai'i (Raiatea). *Journal of the Polynesian Society* 3:136-139.
- Henry, Teuira. 1895. Te parau a Honoura. *Journal of the Polynesian Society* 4:256-294.
- Henry, Teuira. 1897. A war song of the Oropaa clan of Tahiti. *Journal of the Polynesian Society* 6:211-212.
- Henry, Teuira. 1928. *Ancient Tahiti*. Honolulu: Bernice P. Bishop Museum Press.
- Hiro, Henri. n.d. *Pehepehe na te metua vahine*. Tahiti: Henri Hiro.
- Hiro, Henri. 1991. *Henri Hiro*. Papeete, Tahiti: Tupuna Productions.
- Mapuhi, Rui a [Louise Peltzer]. 1985. *Pehepehe - te hia'ai-ao*. Tahiti: Louise Peltzer.
- Mapuhi, Rui a [Louise Peltzer]. 1993. *Pehepehe*. Papeete, Tahiti: Louise Peltzer.
- Pomare II. 1812. Letter to the missionaries, Uaeva, Moorea, October 8, 1812. School of Oriental and African Studies Special Collections: CWM, South Seas Incoming Correspondence, Box 2, 1812-1818.
- Pomare II. 1817a. Letter to Mr. Rowland Hassall, Parramatta, New South Wales, 3 July, 1817. School of Oriental and African Studies Special Collections: CWM, South Seas Incoming Correspondence, Box 2, 1812-1818.
- Pomare II. 1817b. Letter to Revd. S. Marsden, Parramatta, New South Wales, 3 July, 1817. School of Oriental and African Studies Special Collections: CWM, South Seas Incoming Correspondence, Box 2, 1812-1818.
- Pomare II. 1825. *E tura na Tahiti, e Moorea, e na Meetia, Ana, Auura, Matea, e Tetiaroa hoi*. Burder's Point, Tahiti: Windward Mission Press.
- Porinetia: Te 'a'ai e te hiro'a tumu. 2003. *Te Peretitenira'a nō te Porinetia Farani*. <<http://www.presidence.pf/index.php?1762>> [1 December 2005].
- R.T. 1962. A'ai no te mo'o. *Bulletin de la Société des Etudes Océaniques* 138:30-34.
- Raapoto, Turo a. 1990. *Te pina'ina'i o te 'aau, pehepehe*. Papeete, Tahiti: Tupuna Productions.

- Raapoto, Turo a. 2005. Te Muriavai - l'Embouchure. *Direction des Enseignements Secondaires, Polynésie Française*. <<http://www.itereva.pf/grouperlangues/ANNEXES-HTML/poemesdeturo.HTM>> [1 December 2005].
- Te ma'i aho pau. 2003. *Asthma Reality*. <<http://www.asthme-reality.com/tahiti.htm>> [1 December 2005].
- Te mau parau 'āpī: Te pupu vahine maori no te fa'ahorora'a pōpō i te Fare o te Nuna'a. 30 January 2006. *Te Perititenira'a nō te Porinetia Farani*. <<http://www.presidence.pf/index.php?21499>> [5 February 2006].
- Te oraraa taatiraa. n.d. *Service de la Jeunesse et des Sports de la Polynésie Française*. <[http://www.sjs.pf/b\\_sports/Associations\\_Sportives/vie\\_associative/Guide\\_pratique\\_des\\_associations/Tahitien/Mot\\_ministre\\_tahitien.htm](http://www.sjs.pf/b_sports/Associations_Sportives/vie_associative/Guide_pratique_des_associations/Tahitien/Mot_ministre_tahitien.htm)> [1 December 2005].
- Teri'iaama, Vaetua. 2001. Te fare ha'api'ira'a o Ma'atea. *Direction des Enseignements Secondaires, Polynésie Française*. <<http://www.itereva.pf/disciplines/reomaohi/archi/haapiiraa/texte.html>> [1 December 2005].
- TKNui. 12 March 2003a. RE: E aha te faufa'a o te reo mā'ohi i teie tau? [Discussion]. *Fare Vana'a - 'Ōrerora'a* <<http://www.farevanaa.pf/forum/index.php>>. <<http://www.farevanaa.pf/forum/viewtopic.php?p=636&sid=7a3d59f63579390b8f7edfb5a5f78cda>> [1 December 2005].
- TKNui. 12 March 2003b. RE: Te tahu'ara'a natiuira: E aha te faufa'a? [Discussion]. *Fare Vana'a - 'Ōrerora'a* <<http://www.farevanaa.pf/forum/index.php>>. <<http://www.farevanaa.pf/forum/viewtopic.php?t=158&sid=99648acee1311ac58835faa84ff8f85d>> [1 December 2005].
- Yon Yuc, Simako. 1997. Te patiarā'a honu i Ra'ipoto. *Direction des Enseignements Secondaires, Polynésie Française*. <<http://www.itereva.pf/disciplines/reomaohi/archi/rahui/texte.html>> [1 December 2005].

The preceding document was presented at the Tenth International Conference on Austronesian Linguistics (10-ICAL). To properly reference this work, please use the following format:

<LastName>, <FirstName>. 2006. <PaperTitle>. Paper presented at Tenth International Conference on Austronesian Linguistics. 17-20 January 2006. Puerto Princesa City, Palawan, Philippines.  
<http://www.sil.org/asia/philippines/ical/papers.html>

For other papers that were presented at 10-ICAL, please visit <http://www.sil.org/asia/philippines/ical/papers.html>.