

英語俳句データベース構築の試み

木村映善¹、墨岡学²、田中喜美代³、和田武¹

俳句 ML の参加者の意識から構成される俳句の意識空間 (= Haikusphere) を単語と句読点という構成要素からなる集合として定義し、そこから俳句の形式議論を展開しようという試みを始めている。しかし、非英語圏俳句は構文解析等の技術的困難さから、抽出が比較的容易と思われる英語俳句を選択した。3行形式が主流になっていること、メールのエンコーディング傾向より英語圏に所属するネイティブによる投稿が多いと思われる、Shiki Team が主宰している Nobo メールリストをサンプリング対象とし、それらのメールからの俳句の抽出とデータベースの構築を試みた。

An approach to constructing Haiku relational database by the Shiki team

Eizen Kimura, Manabu Sumioka, Kimiyo Tanaka, Takeshi Wada

Our approach to constructing Haiku relational database model is based on the archives of Haiku mailing list Nobo. English Haiku as a short poem is composed with three lines. We used "Message-ID" field as to define "Haiku ID" and "References" field to relate and trace the original Haiku. We already have more than one hundred thousand words from the Nobo mailing list spool. Soon, web access to the haiku database server will be available via MySQL.

1. はじめに

本論は、英語の俳句(Haiku)のデータベース化を、誰に取っても解り易く、また利用し易いような形式で提供するために、我々が行おうとしている英語俳句データベースのひとつの試みを論じたものである。基本となるデータベースの対象は、我々がこれまで Web で公開してきたメールリスト(ML)をもとにしている。

Nobo ML への 10360 通(俳句以外もあり)の投稿について、HTML メール対応メーラー(ちなみに例を上げれば、Nobo の参加者の中での Microsoft Outlook の利用率は、X-Mailer:項を見ると 56%になる)普及に伴い、Multipart Alternative で通常のテキストと HTML 形式で投稿してくるものがほとんどである。俳句投稿者の地理的分布の類推を調査するために、メールの charset 属性を調べると、次のようになった(表1)。最も多いのは ISO_8859-1 が 50.6%で、ISO_8859-2 はごく少数で 7 通にすぎない。また US-ASCII も 42.9%で半分近くを占めている。もともとは、US-ASCII(ANSI_X3.4-1968)が主流で HTML メール対応メーラーの普及にともなって ISO_8859 が増加傾向にある。メールユーザエージェント(MUA)を調べるとさらに詳しいことがわかるが、こ

¹ 愛媛大学総合情報メディアセンター(Center for Information Technology, Ehime University)

² 松山大学経営学部 (Faculty of Business Administration, Matsuyama University)

³ 愛媛大学留学生センター (International Student Center, Ehime University)

では述べない。正確には、1994 年からの Shiki ML を調査すると時代とともに変化する charset の傾向はさらに明らかになってゆくはずであるが、ここでは、この問題にはこれ以上深入りせず、英語文化圏を対象とした Haiku の構造解析とデータベース構築を行う。

ISO_8859-1,2	50.6%
US-ASCII	42.9%
Windows-1252	4.9%
UTF-8	1.5%
Other	0.1%

表 1. Nobo ML の文字コード

WHChaikuforum@yahoogroups.com
bashos-beatniks@yahoogroups.com
nobo@shiki1.cc.ehime-u.ac.jp
WHCworkshop@yahoogroups.com
Dragon_haijin@yahoogroups.com

表 2. 世界の Haiku ML の例(一部)

メーリングリストの参加者は、約 300 名である。ほとんどの人がフリーメールでの参加で、そのために国籍などは不明だが、95%以上が、米国、ヨーロッパで Haiku を創作する人々と推測される。この推測の根拠のひとつには、Nobo ML への投稿のほとんどが日本時間の午前 0 時を過ぎてからで、日本人の昼間の仕事時間にはほとんど投稿がないことがある。ただし、日本に住んでいて昼間仕事をして、深夜 Haiku 創作活動をしているのではないかと疑いもあるが、現実に本人を知っている Haijin 達やメールヘッダから読み取れるアクセスプロバイダの時刻などから判定しても、Nobo ML への参加者は、米国やヨーロッパで Haiku 創作活動をしていると判定される。この事実もメールで利用されている文字コードが US-ASCII 中心であることを示している。

議論の対象は、基本的に US-ASCII 文字を中心とした Haiku についてのデータベースである。それ以外のアジアの文字による Haiku(俳句)は今のところ除いている。しかし、韓国での Haiku 活動なども一部盛んな動きもあって将来、アジアで我々のようなオープンな Haiku のメーリングリスト活動が始まれば対象に入れることも必要になってくるため、それらを意識的に除外したものではない。我々が作成したものを中心として考えようということである。特に、日本語を除外しているのも特別視したものではなく、日本語俳句も取り上げるメディアと愛好者の広がりを考えれば「難しいから」と放って置くわけにはいかない。特に、携帯電話サイトでの俳句については、伝統俳句協会の事務局長とも個人的にお話をしたこともあり、2000 年には、短期間であったが我々は四国アステルとの共同で PHS での実験的な「ハイクマダーム」を試みたこともあった。四国アステルのメニューのひとつとして、自由なケータイ俳句の投稿を受付し、シキチームのひとりがコメントをつけて、アステルのメニューの中で閲覧できるようにしたものである。これは四国地区限定だったが、参加者が少なく自然消滅の形になってしまった。もっと昔には PDA の先駆けアップルの Newton が出たときには即買い込み、これで携帯俳句端末を、と意気込んだが日本語の縦書き表示に苦しんだ経験もしてきた。

そこで Haiku の対象を ASCII 文字のものに限定した。現在、ネットの世界を見回してみると、アクティブな ASCII 文字を中心とした Haiku のメーリングリストには 180 サイト以上が存在し、表 2 にその一部を示す。Yahoo! の e-group 機能を利用した yahoogroups.com ドメインのメーリングリストはそのグループのメンバーだけの閉じたものである。グローバルにオープンにしている Haiku メーリングリストで最大のものは、我々の運用している Nobo であろう。このことは、運用についての責任の重さと同時に、玉石混交ではあるが最大のアーカイブを保存している我々のサイトがこれからやらなければいけない仕事は、サーバに積み残されたまま 20 世紀の Shiki から 21 世紀の Nobo までのオープンな Haiku サイトに蓄積された投稿を整理し、Haiku 初心者から Haiku 研究者にまでデジタルアーカイブをデータベースとして提供することではないかと考えている。ちなみに、googlebot による Shiki Haikusphere へのアクセスは、開設した 2003 年 1 月 1 日から今までの 1 日当りの平均は約 3300 回であった。これまで、データベースをつくるために必要な腕力を持った人材がいなかったが、第 62 回人文科学とコンピュータ研究会発表会での我々の発表後、若手の協力者を発掘し、今回新たにデータベースを作成する試みを開始した。

メーリングリストへの投稿は、Haiku、Haibun(俳文)、Haiga(俳画)が主体である。ただ、それ以外の Haiku コンテストの案内、Haiku についての質問などもある。メーリングリストへの投稿から Haiku データベースをつくるために、自動で Haiku を抽出するプログラムの完成を目指して

いる。Haiku かどうかの判定としては、3 行で固められたものという、素朴で形式的なものを使う。このために、どうしても雑音、ゴミが切捨てられない。ただ、それらは、一度人間の眼で見ることによって、簡単なフィルタをつくることができる。このフィルタは、しだいに精度を上げていくことが可能で、ちょうど Mozilla が採用しているスパムフィルターのアルゴリズムのようにページアン理論を応用することも将来は可能であろう。また、根本的な問題、玉石混交の投稿 Haiku のなかからあまりに酷いものを篩い分けることも必要になってくる。メーリングリストへの参加者は、酷い投稿をする人のメールアドレスによって、これを自分のコンピュータの MUA でやっているが、それをメーリングリスト運用者がメールアドレスによって差別をするのはいかなものかと思う。我々の Old Shiki や New Nobo は、いわゆる No Sensei のルールで運用しているメーリングリストであるのでリスト運用者が管理者になって篩い分けをしてデータベースをつくることは、もともとの精神に反する。また、良い Haiku とは何か？という質問に答えることは、これも難しく、良い俳句を選択することも容易ではない。では、ある程度以上のレベルと判定することは、プログラムで可能か？という質問も難しい。Haiku の定型性として日本語の音のリズム 5-7-5 に相当する 3 行詩(L1,L2,L3)としての表現形式をゆるやかな定型性として今回のデータベースは作られた。ちなみに 1950 年代の古い Haiku は、先頭大文字ではじまり必ずピリオドで終わっている。

2. Haiku と Haikusphere 集合の性質

Haikusphere という造語を作り、Shiki Haikusphere と名付けた web サイトを作って、そこで Nobo と名付けた主として Haiku、Haibun、Haiga のためのメーリングリストをはじめた。ここに集積される Haiku データは、nobo@haiku.cc.ehime-u.ac.jp に属するものである。Old Shiki メーリングリストは、shiki@cc.matsuyama-u.ac.jp に属するものである。それぞれの Haiku 集合はそれぞれの Haikusphere に属している。Haikusphere は、メーリングリスト参加者の Haiku についての意識が作り出した空間と直感的に理解される。Old Shiki の場合、インターネットでの最初の Haiku メーリングリストであったために、さきほど述べたゆるやかな定型性が確立していなかった。One liner と呼んだ Haiku も少なくない。機械的な Haiku の抽出をメーリングリストへの投稿から行うために 3 行詩の形式がほぼ参加者の意識に浸透した Nobo を対象にデータベースを作ることにした。Haikusphere のモデルを解説しよう。図 1 は、Nobo ML への実際の投稿例である。

図 1. Nobo への俳句投稿とコメントの例

Date: Mon, 21 Jun 2004 02:00:43 -0400	moon shadows
From: Paulafish@aol.com	on white roses
Reply-To: nobo@shiki1.cc.ehime-u.ac.jp	a cricket sings
To: nobo@shiki1.cc.ehime-u.ac.jp	-----
Subject: (Nobo:10536) Re: moon shadows	Rose Marie Stutts c&c welcome
MessageId:<07320285.04686DC2.0272AF5D@aol.com>	Rose Marie, I think it's beautiful.
	Thank you. paula

Nobo への投稿は通常、「Haiku+作者名」で行われる。さらにコメントや批評を欲しい場合、'c&c welcome' や 'c/c welcome' などと略記して追記する。'c&c(c/c)' は、'comment and (or) criticism' の略である。他人からのコメントや批評などは欲しくない場合もあり、その場合もはっきりと 'no c&c' と明記する。不毛な論争をさけるためのアイデアで、ほぼ Nobo の暗黙のルールになっている。Web 日記サイトなどでのコメント無効/有効のスイッチに相当する(例えば、slashdot.jp)。

はじめに Haiku ID の付け方の例を説明する。図 1 は、Rose Marie Stutts へ paula が haiku のコメントを送ったものである。Stutts が投稿したオリジナルの Haiku は Nobo:10535 にあり、

それは'moon shadows/on white roses/a cricket sing'である。Nobo の ML スプールから haiku を抽出するためには、この Nobo:10535 の haiku に ID を割り当てて抽出しなければならない。Nobo:10536 の haiku は、Nobo:10535 の haiku を参照/引用したものである。プログラムでデータベース構築のために Haiku を自動抽出するときに気をつけなければいけないのは、paula は、Rose Marie Stutts の haiku を引用符なしで丸ごと引用していることである。Nobo:10536 から抽出された Haiku は、paula のものではなく、Rose Marie Stutts のものである。このために Stutts のメールの"Message-ID"がこの haiku の Haiku ID となる。

図 2. Haiku ID ("Message-ID")と Haiku 本体

Nobo:10535 の"Message-ID:"項	-> 'moon shadows/on white roses/a cricket sing'
Nobo:10536 の"Message-ID:"項	-> 'moon shadows/on white roses/a cricket sing'の引用

図 3. Haiku の参照("References")

A: "References:"項の"Message-ID:"項	-> Haiku 引用元のメッセージ
B: "References:"項がなく、前出 Haiku(引用)がある	-> Nobo サーバの受付時刻順

通常の MUA で References 機能を利用すれば、Nobo:10536 の"References:"項に引用元の"Message-ID:"項の内容が埋め込まれるためにメッセージのリンクができ、オリジナルの Haiku が正確に検索できる。ただし、この Nobo:10536 のように References 機能を用いず、送信者のコンピュータ内部でコピー・ペーストを行った場合は、Nobo サーバが受け取ったメールの時刻順に早く出現したものがオリジナル、あとは引用として Haiku の自動抽出では処理をするしかない。

"References:"項がメールヘッダに無い場合でも、次の図 4 のような例では、メッセージ本文中に引用の見出しがあるためにオリジナルの Haiku を関係付けることができる。この図 4 は、投稿者のメールプログラムがメッセージ本体には引用の見出しをつけているにも関わらず、メールヘッダに"References:"を付けていない例である。

図 4. Nobo への投稿、オリジナルの引用見出しがある場合

Date: Mon, 21 Jun 2004 11:26:02 EDT	moon shadows
From: Maleti@aol.com	on white roses
Reply-To: nobo@shiki1.cc.ehime-u.ac.jp	a cricket sings
To: nobo@shiki1.cc.ehime-u.ac.jp	-----
Subject: (Nobo:10537) Re: moon shadows	Rose Marie Stutts
	c&c welcome
In a message dated 6/21/2004 12:30:37 AM	
US Eastern Standard Time, rmstutts@ bama.	Beatiful, Marie!
ua.edu writes:	Mary Lee

この図 4 のような例では、メッセージ本文中の引用見出しを切り出して処理をしてもいいが、図 3 の B のような処理をすることにする。データベースの基本となるモデル作成にあたり、どこまで豊富な機能を入れるかではなく、どれだけ捨てて身軽になれるかを方針に立てる。単純なモデルとして Haikusphere を組み立てていく。

Nobo Haikusphere は、Nobo ML に投稿された Haiku によって構築されている。構築は、日々、参加者の Haiku 投稿とコメントや批評によって行われる。Nobo Haikusphere を構成する基本となる要素は、Haiku そのものである。印刷物としての英語 Haiku については 50 年以上の研究があるが、メーリングリストのようなネットの世界での Haiku も同じように取り扱ってもよいのであろうか？メーリングリストの多くの参加者によって参照/引用された Haiku はそれだけ多くの人の共感を呼んだことになり評価は高いということになっているが、良い Haiku、つまらない Haiku をネットのメーリングリストを通して学習することができるのであろうか。

ひとつの実験として、日本の学生を使って Nobo に英語俳句を投稿させてみた。この学生達は、

特に英語が得意でもなく、普段英語を書いたり話したりしている学生でもない。また、正岡子規記念博物館がある街に在住の学生であるが、日常的に俳句に触れることはない学生達を選んだ。当然のように、彼らの投稿に対して Nobo のメンバーからなんの反応もなかった。彼らの Haiku は無視されたのであろうか？

ところが、しばらくリストの様子を見てみると、ひとりの学生の Haiku に、その引用とそれを改良した、あるいは発展させたような Haiku が或る米国人(A)によって投稿された。この米国人は、我々が主催したシキハイクコンテストで優勝した方であった。さらに、もうひとりの学生の Haiku が引用され、連句のようなものが投稿された。これは別の或る米国人(B)からだった。学生達には、まず日本語の俳句を作ること、その後で、それを自分なりに英訳するように、とだけしか指示をしていなかった。最初から英語の単語を並べて作るのではなく、曲がりなりにも日本語でしっかりと考えて俳句を作り(ただし、日本語俳句の訓練なしで)、その後で翻訳を試みよ、と自分の頭で考えさせた。ところが、筆者はすっかりだまされていた。彼らは、機械翻訳サイトを使って日本語の俳句を英訳してしまい、それをそのまま投稿したと言うのである。

しかし、ハイクコンテストで優勝経験のある米国人 A から反応があった Haiku の元の日本語の俳句を作ったのは、その学生の母親で或る俳句結社で活動している人であったことに驚いた。機械翻訳は英語の Haiku を作れるほど進化しているのであろうか？このことがわかった後で、学生に直接米国人 A へ手紙を書かせた。内容はお礼である。A からの返信には A 自身は日本語の俳句を理解する能力がないこと、学生には英語の力は弱いけど、共感があったというようなことが書かれていた。もうひとつの例、米国人 B からの反応があった学生の場合も非常に興味深い例であった。この学生は、いわゆるゲームオタクですっかりコンピュータゲームの世界にはまっている学生であった。haiku の内容は、ゲームについてはなかったのだが、反応してきた米国人 B がたまたま自分の Web サイトの URL を公開していたのでそのサイトを見ると、驚くべきことにサイバーオタクとでも呼べるものであった。ただし、B は自作の Haiku(のようなもの)を 1 日一句のような形式で Web に書き込んでいるのでネットで Haiku の達人を目指しているようだ。

この実験に参加したのは、7 人の学生で短期間であったため、一般的な結論を出すことはできなかった。また得られたこの奇妙な結果も偶然の結果とも判定される。しかし、英語 Haiku ML を 10 年間運営してきて、何かメーリングリストの方向性のようなものが偶然の中に見られることがしばしばあった。混乱したとき、モデレータは介入を最小限にして、メーリングリストの意思のようなものにまかせることが混乱を収束させる、といった経験を何度もしている。

そこで、Haikusphere はメーリングリストごとに実在するとの仮説を立てる。Haikusphere は、Haiku を要素として構成されるとする。ここで Haiku を 3 行詩と仮定し $\langle L1/L2/L3 \rangle$ のように表記する。まず、もっともつまらない Haiku としては、空の投稿、つまり空 $\langle \rangle$ がある。さらに意味の無いアルファベット一文字で構成される Haiku が続く。例えば、実際に $\langle a/b/c \rangle$ のような投稿があった。投稿は、US-ASCII の文字集合で構成されるとすると、その文字集合の要素であるアルファベットからなる単語と空白で L_i ($i=1,2,3$) は構成される。また、ダッシュやコロンなどの切字 (cut) $\{-, :, \dots\}$ は行 L_i を構成する最後の単語の後ろに付く。3 行詩の改行を表す文字 \backslash も、一種の切字と解釈することも可能である。

Haiku: $\langle L1/L2/L3 \rangle$ が 3 行詩であるためには、 $L1, L2, L3$ がすべて空でない必要がある。 $\langle \rangle$ は空の 3 行詩だが、現実には取り扱っていない。単純に Haiku を構成する各行 $L1, L2, L3$ のそれぞれの単語数と切字の数をあわせて n として、 H_n について順序関係を作ることができる。

$$H_n \quad H_m \quad \text{if } n \leq m.$$

集合 Haiku_N の任意の要素 $H_p, H_q, H_r \in \text{Haiku}_N$ について、次のような半順序関係が成り立つ。

- (1) $H_p \leq H_p$
- (2) $H_p \leq H_q \leq H_r$ ならば $H_p \leq H_r$
- (3) $H_p \leq H_q \leq H_p$ ならば $H_p = H_q$

単語数と切字の数の合計だけで構成される集合 Haiku_N について、 $n=0,1,2$ や $20 \leq n$ などは、Haiku としては意味が無いことが自明である。実際に Nobo ML から Haiku_N を作っても興味深い性質は発見されない。具体的に英語俳句で使用されている単語をもとに集合を作りその性質を調べることにする。

Nobo Haikusphere を、Nobo サーバのプールファイルから抽出された Haiku の集合と定義し、Haikusphere(Nobo)と表記する。この集合の性質がどのようなものか。Haikusphere(Nobo)の構造が英語俳句の一般的なものなのかどうか。さらにネットで公開されている各種の Haiku メーリングリスト X についてその Haiku 集合を Haikusphere(X)としたとき、Haikusphere 間の関係を調べるのが今後の課題である。

Haikusphere (Nobo): Haikusphere(X)

3 . データベース構築の背景

「E-zine 発行者が見た英語俳句の形式議論」(文献 1)ではライクホールド氏による良き Haiku を作る為のルールブックの中から「切れ」に使われている句読点が複数あることを紹介している。英語俳句、すなわち Haiku は現在、3 行構成が主流で、「切れ」の為に使われる句読点は L1 か L2 の最後に配置されることになっており、その句読点については図 5 のように意味の含有があることが示されている。

図 5 . 句読点

「:」 停止	「,」 軽い休止
「;」 半停止、あるいは休止	「--」 同じことを他の言葉で言い換え
「...」 述べられていないこと	「.」 完全停止

また、「切れ」は 3 行もしくは 2 行形式で表現するときの改行の入れどころによっても表現される。しかし、俳句では「切れ」は「かな」「や」「けり」「らむ」など代表的なもの 18 個があることが指摘されており、詠嘆、省略、休止、強調、格調、提示、願望、疑問、並列、反疑、命令といった働きを持ち、かつその働きが単独のものとは限らず、複数の働きを併せ持つという重層的構造を持つ。そのため、俳句の多様な「切れ」の表現様式とその重層的構造を Haiku の様式に翻訳する時に本家の俳句以上に制約されたルールの中に落とし込まなければならず、大元の俳句にある意味をどれだけ落とさずに伝えられるかどうか、翻訳者が苦勞する部分となっている。そこで、Haiku では「切れ」の表現様式の少なさを句読点の複数配置と空間的配置の組み合わせによりカバーしているのではないかという仮説を立てた。しかしながら、2,3 行の中に繰り返し句読点を用いるというくどさは情念、時間的空間に広がりをもたらす一方で Haiku の簡索性、洗練性を失わせる危険性を持つことが考えられる。その仮説を裏付ける為に Nobo-ML(*1)から複数の句読点を用いている Haiku を抜き出して検討した。その結果、下記に例示するようにいくつかの興味深い現象が観察された。

- ・「:」は他の句読点との併用を排除する力がある。
- ・ライクホールド氏は「--」「---」は「同じことを他の言葉で言い換え」と説明しているが、そうとも言い切れない使い方の方が目立つ。

このことから複数の句読点と配置はライクホールド氏の解説にて取り上げられている句読点の一義の意味だけではなく、配置する場所や併用される句読点との相互干渉によっても微妙な意味合いの変化を取っていることが伺える。しかし、この段階では手動で無作為に集めたものであって仮説の域を出ず、また発見されていない用法があるものと思われる。そこで多数の Haiku を ML から収集し、データベースを構築することにより、これらの問題意識に対しての分析をする為の土台を提供することにした。

4 . データベースの設計

データベースを設計するにあたり、以下の点を念頭において設計した。すなわち、(1) Haiku

の構造を表現すること、(2) Haiku 及び作者のユニーク性を確保すること、(3) Haiku 同士の関係性を表現すること、である。

(1) Haiku の構造の表現

Haiku は大体において最大3行より構成されており、各行に数個の単語及び句読点がならぶ形になっている。そこで、Haiku の構造を表現するのに、行・位置・Word の3要素で表現する。例えば図6の俳句(Nobo-ML より引用)は、図7のように表現されることになる。

図6 . 俳句例

Panic of the crowds; summer sky of coiling light- there is no return
--

図7 . 構造例

L1(1) : Panic L1(2) : of L1(3): the L1(4) crowds L1(5) ; L2(1): summer L2(2): sky L2(3):of L2(4):coiling L2(5): light L2(6): - L3(1): there L3(2): is L3(3): no L3(4): return

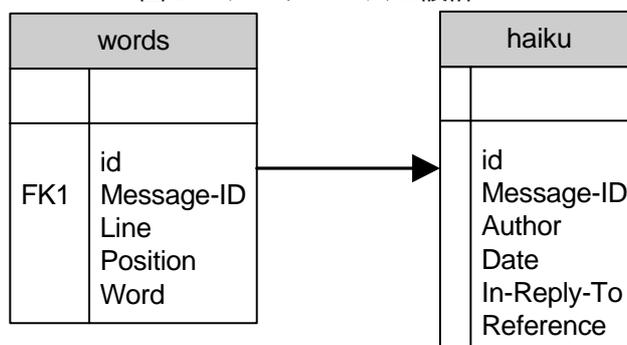
(2) Haiku を一意に同定するためのユニークな属性について

Haiku をデータベース上で扱うとき、それぞれの Haiku を一意に同定する為のユニークな ID を持たせる必要がある。その理由は、いくつかある。一つは前述した Haiku の構造をなす構成要素群がどの Haiku に所属するかを指定するためである。もう一つは Haiku と作者を関連付けるためのキーである。そして最後の一つは Haiku 同士の関係、すなわち連句、返句、参照といったものを表現するために参照する Haiku を一意に指定するためである。現在、俳句及び Haiku を電子的に表現する統一スキーマは定められておらず、XML スキーマの定義など今後の課題になっているが、当座の問題として少なくともメーリングリストに流れている Haiku に関してその一意性を表現出来る必要があった。そこでメーリングリストに投稿されている一通のメールと Haiku は基本的に1対1の対応をとるものとみなし、そのメールにおいてユニーク性を保障するための情報を Message-ID から取ることにした。Message-ID とは RFC2822 “Internet Message Format”(文献2)に定められている、メールの一意性を表現する為のヘッダ情報である。このヘッダ情報をデータベース上で Haiku を特定する為のキーとして利用することにした。

(3) Haiku 同士の関係性の表現

Haiku は投稿された Haiku 単体で完結することもあるが、連句、返句、参照といった関係を持った形で広がりを持つこともある。また作者が複数の俳句を投稿している場合もある。いずれも、Haiku の構成要素の位置関係より上位の次元に位置する関係であり、Haiku 単体の情報とは別に管理する必要がある。Haiku の一意同定は先述した通り、Message-ID で表現されており、またメールにおいて参照、返答が References:、In-Reply-To ヘッダで参照先のメールの Message-ID が記載されている。これを利用すればメールと Haiku の対応は1対1という前提の元に Haiku 同士の関係性を表現することが出来る。さらに、付随情報として、作者(投稿者)名の代わりに From ヘッダから採取したメールアドレス、投稿日時として Date ヘッダから採取した。これらを考慮した結果のデータベース設計を図8に掲載する。

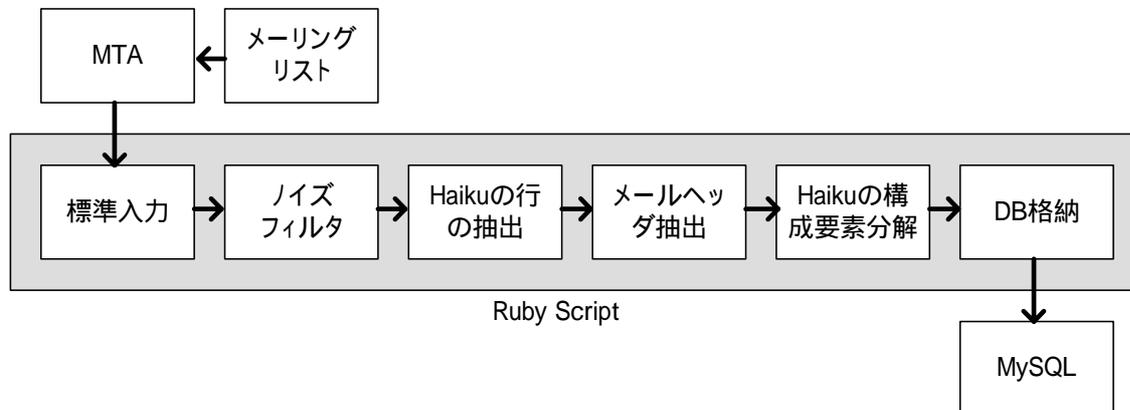
図8 . データベースの設計



5 . Haiku フィルタ及びデータベースの実装

メールからの Haiku の抽出には Ruby を、格納するデータベースには MySQL を利用している。稼働しているメーリングリストからの自動的な取り込みを実現することを想定して、図9に示すような実装を行った。

図9 . Haiku、データベースの実装



Nobo の ML が開設された 2003 年 1 月から現在 2004 年 5 月までのメールから抽出した結果、5660 句の投稿があり、それらの Haiku を構成している単語の総数は 103200 個であった。以後、これらのデータを元に様々な分析が進むことが期待される。

6 . まとめ

Haikusphere に属する Haiku 集合の現実の構造を調べるための枠組みを作り、Haiku の創作や研究に役立つネットの中での仕組み作り出すのが今回のデータベースの試みである。例えば、一般に英語俳句では季語(season word)は不要というのが定説のようにになっている。しかし、英語の歳時記を作る試みをしている Haiku 研究者もいる。このために、今回のデータベース構築の試みは、この歳時記のような試みにも役立つものと願っている。

参考文献

- 1 墨岡,田中,和田, E-zine 発行者が見た英語俳句の形式議論, 2004-CH-62, pp21-28,2004.
- 2 “Internet Message Format” RFC2822 P. Resnick, Editor, Network Working Group April 2001.
- 3 H.G.Henderson, 'An Introduction To HAIKU', Doubleday, 1958.
- 4 R.H.Blyth, 'HAIKU Volume 1 Eastern Culture', The Hokuseido press, 1949
- 5 Shiki team, '俳句@インターネット', シキチーム出版、2000 年
- 6 Shiki team, 'The Shiki Internet Haiku Salon', シキチーム出版、1997 年
- 7 Shiki team, 'Shiki team からのメッセージ', シキチーム出版、1999 年
- 8 墨岡, 田中, 別宮, ライクホールド,'俳句の「場」としてのインターネット、松山大学総合研究所、2003 年.