

国立国語研究所学術情報リポジトリ

<全文>Evidence-based Linguistics Workshop 2022 発表論文集

メタデータ	言語: jpn 出版者: 公開日: 2022-10-21 キーワード (Ja): キーワード (En): 作成者: 国立国語研究所 メールアドレス: 所属:
URL	https://doi.org/10.15084/00003670

This work is licensed under a Creative Commons Attribution 3.0 International License.



Evidence-based Linguistics Workshop 2022

発表論文集

2022年9月5・6日(月・火)

大学共同利用機関法人 人間文化研究機構
国立国語研究所 編

目次

- ChaKi.NET lite の開発: Universal Dependencies コーパスの利用を見据えた ChaKi.NET
ユーザインターフェースの改良
伊藤薫 (九州大学)・森田敏生 (総和技研) 1
- アタヤル語群における「言う」の再建
落合いずみ (帯広畜産大学) 6
- 編集後記
浅原正幸 (国立国語研究所) 17

ChaKi.NET lite の開発: Universal Dependencies コーパスの利用を見据えた ChaKi.NET ユーザーインターフェイスの改良

伊藤 薫 (九州大学) †

森田 敏生 (総和技研)

Development of ChaKi.NET lite: Improvement of the User Interface of ChaKi.NET for the Universal Dependencies Treebank

Kaoru ITO (Kyushu University)

Toshio MORITA (Sowa Research Co., Ltd.)

要旨・既発表の有無

ChaKi.NET はコンコーダンスやアノテーションツールを含む多機能なコーパス管理システムである。現行の (ChaKi Legacy でない) ChaKi.NET は 2009 年 (Ver. 1.1, 現在公式サイトで確認できる最古のバージョン) に公開されたものであり、10 年以上に渡って機能追加や品質の向上が行われてきた。しかし、可能な処理が増え多機能化が進む一方で、インターフェイスが複雑化しコンコーダンスとしての利用等、簡易な作業にも学習コストが高む状態になっている。本研究では、ChaKi.NET の機能のうちコンコーダンス機能に焦点を絞り、複雑なファイル操作に馴染みのないユーザにも利用しやすいインターフェイスを備えた ChaKi.NET lite の開発について述べる。想定するユーザ層としては Corpus of Contemporary American English (COCA) 等のコーパスを使用したことのある言語学者であり、近年自然言語処理分野で開発が盛んである通言語コーパス群 Universal Dependencies (UD) コーパスを容易に使用可能にすることを目指している。今回追加した主な機能は、UD コーパス群読み込み操作の簡略化、複数 UD コーパスの一括検索機能追加、検索実行および結果表示インターフェイスの改良である。

既発表無。

1. はじめに

ChaKi.NET (Matsumoto et al. 2006) は多機能なコーパス管理システムであり、コンコーダンスやアノテーションツールなど、自然言語処理向けのアノテーションやデータ閲覧にとどまらず、言語学の研究にも有用なツールである。ChaKi.NET の持つ機能のうち、言語学研究において今後有用だと思われるのが様々なフォーマットのコーパス読み込み機能や、タグや係り受けなど、様々な統語情報を指定可能な検索機能が挙げられる。現在主に用いられている ChaKi.NET では、2009 年に Ver. 1.1 として公開されたものに様々な機能が追加されており、様々な処理が可能であるが、機能の豊富さや歴史の長さ故に新規ユーザにとっては学習コスト

† ito@flc.kyushu-u.ac.jp

が高く利用しづらい面も抱えている。

そこで本研究では、ChaKi.NET の機能をコンコーダンス関連に特化し学習コストを下げることにより、気軽に利用可能なツールである ChaKi.NET lite を開発する。ChaKi.NET lite デザインの方針は下記の通りである。このうち、最後の Universal Dependencies (UD) の重要性については次節で述べる。

- インターフェイスの改良により直感的に利用可能であること
- ChaKi.NET のデータベースと互換性があること
- 言語学研究に適した機能を持つこと
- コーパスを利用して言語学研究を行うユーザを主な対象とすること
- UD ツリーバンクの操作に適していること

なお、本論文のスクリーンショットは全て開発中のものであり、実際に公開するツールとは仕様が異なる場合がある。以下、本論文では UD を重視する理由について述べた後、主要な改良点を紹介する。

2. Universal Dependencies (UD)

UD は通言語的に統一された方法で依存構造や品詞などについてアノテーションが付与されたコーパス群を開発する国際プロジェクトである (Björkelund et al. 2017)。UD は依存構造が内容語主辞で付与されていること、Universal POS (UPOS) と呼ばれる品詞体系や依存関係タグが定義され、コーパスに付与されていることなどを特徴とする。また、人手アノテーションのしやすさやパーズング精度の高さなど、言語学的妥当性以外も考慮して設計されている (浅原ほか 2019)。UD ツリーバンクの各コーパスは CoNLL-U というフォーマットで書かれているが、この形式は人間が直接依存構造木を読むことや、例文を検索することには適していない。本研究では、CoNLL-U の読み込みやタグ、係り受け情報の検索、依存構造木表示など、UD コーパスの利用に適した ChaKi.NET を改良することで、言語学や関連分野の研究を促進するツールの開発を目指す。

言語学研究における UD ツリーバンクの利点は、主にデータの多様さにある。最新バージョンの Ver. 2.10 では 130 言語、228 ツリーバンクを収録している。これにはチュクチ語やワルピリ語などの少数言語、アッカド語やコプト語、古代中国語 (漢文) 等の古代語、ヒンディー語—英語やトルコ語—ドイツ語などのコードスイッチング、学習者コーパス、手話 (スウェーデン手話) などが含まれ、個々のデータ量は少ない場合もあるが時空間的、様態的に非常に幅広い言語が収録されている。また、先に述べたように UD は同じ枠組みでデータが記述されているため横断的な検索が容易である。つまり、品詞を例に挙げると UD では UPOS と呼ばれる統一された品詞体系でアノテーションされているため、一般名詞と接置詞 (前置詞・後置詞) の係り受けなどを検索したい場合は全ての言語で NOUN タグと ADP タグの係り受けを検索すればよく、個別の言語に合わせてクエリを記述する煩雑さが軽減される。また、多くのコーパスは XPOS と呼ばれる個別言語に特化した (コンバージョン元コーパスの情報であることが多い) 品詞情報を保持しており、個別言語に特化した検索も可能である。また、多くのデータがフリーで利用できることも、多言語データへの気軽なアクセスを可能にしている。このような

特徴から、UD ツリーバンク利用促進は類型論研究に役立つことが期待される。

既存ツールとしては UD 公式サイトで SETS treebank search, PML Tree Query, Kontext, Grew-match, INESS の 5 つが紹介されているが、これらはいずれもそれぞれ強みを持つものの、検索クエリ入力の記述方法が複雑であり、記法を習得するのに学習コストがかかる。本研究では、もともと GUI による検索クエリ記述が可能である ChaKi.NET を UD 向けに改良し、より簡便な検索を可能にする。

3. インターフェイスの改良

本節では、ChaKi.NET からの主な改良点について述べる。主な改良点は画面全体に表示される要素とその配置、ドラッグ&ドロップによるツリーバンク読み込み機能の追加、複数コーパスの表示改良および選択機能、操作フローの改善であり、これらを順に紹介する。

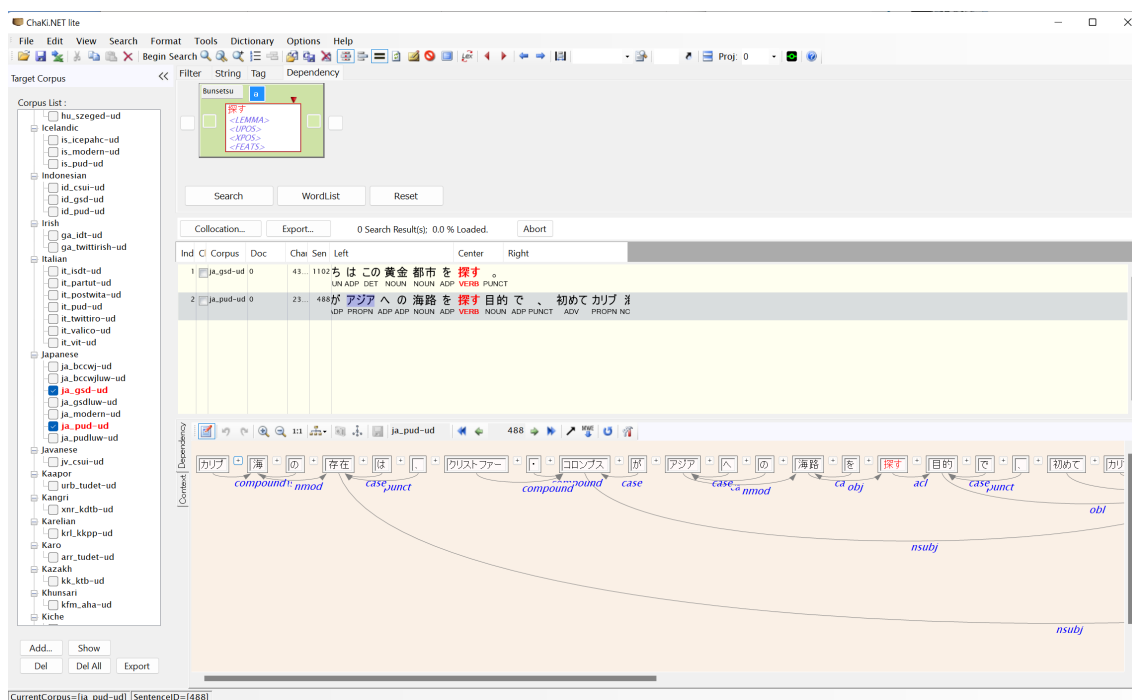


図 1 ChaKi.NET lite (開発中) の画面

3.1 基本設計

ChaKi.NET にはアノテーション機能が含まれるが、言語学の研究においてはアノテーション機能を用いる機会は少ない。実際、主要なコーパスである Corpus of Contemporary American English (COCA), British National Corpus (BNC), 『現代日本語書き言葉均衡コーパス』(BCCWJ) にも主要なインターフェイスにアノテーション機能はなく、需要も少ないと思われる。そこで、ChaKi.NET lite では需要の高いコンコーダンス機能に特化し、1 画面で操作が完結することを目標に設計した。具体的にはコーパス読み込み、係り受け検索、結果表示、係り受けツリー可視化という一連の操作を直感的に行えることを目指し、ChaKi.NET

から表示する機能やペインの絞り込みを行った。これを図 1 に示す。図 1 の左列でコーパスリストの読み込みやコーパス選択、右列上段で係り受け検索、右列中段で検索結果表示とツリー表示対象文の選択、右列下段で選択した文のツリー表示を行うことができる。

3.2 ドラッグ&ドロップによる UD ツリーバンク一括読み込み

ChaKi.NET では.conllu ファイルを読み込む際、アプリケーション上の機能を用いて.db ファイルに変換する操作が必要であった。この操作は『中納言』を介した BCCWJ の利用、あるいは、ウェブ上のインターフェイスが提供されている BNC, COCA を始めとしたコーパスを中心に利用している研究者にとっては使用上の障壁になる。そこで、UD ツリーバンクデータから事前に変換・提供される.db ファイルの格納されたフォルダを画面上にドラッグ&ドロップすることで一括読み込みする機能を追加した。これにより、複雑な前処理なしで大規模な通言語コーパスの利用を可能にした。

3.3 複数コーパス選択機能

ChaKi.NET lite では 3.2 で述べたツリーバンク一括読み込みの後、収録コーパスを複数選択し、横断的な検索が可能である。コーパスは UD ツリーバンク内のフォルダ構造を反映し、日本語、英語など言語ごとにまとめて表示され、クリックにより折りたたみや展開が可能である。また、複数選択に関しては各コーパス横のチェックボックスをクリックすることで検索対象とするか否かを選択でき、選択されたコーパスは赤字で強調表示される。

3.4 操作フローおよび結果表示の改良

既存の ChaKi.NET では、検索クエリ入力画面と検索実行ボタンが離れた場所に配置されており、検索クエリの入力後に検索を実行する手順が直感的にわかりにくい状態であった。ChaKi.NET lite では Google 検索や COCA など、多くの想定ユーザが慣れ親しんでいると思われるインターフェイスに合わせ、検索クエリ入力場所の直下に検索実行ボタンや語彙リスト表示ボタン、検索条件リセットボタンを配置することで、直感的な操作を可能にした。他にも検索ヒット数の表示箇所をコンパクトにするなど、検索クエリ入力から例文表示、ツリー表示という一連の操作に集中できるようなデザインにした。

4. おわりに

本論文では、現在開発中の ChaKi.NET lite について、開発目的、主な改良点と想定ユーザを中心に述べた。派生元の ChaKi.NET が総合的なコーパス管理・編集・検索システムであるのに対し、本研究で開発した ChaKi.NET lite は、インターフェイスの改良とコンコーダンスとして機能を絞り込むことで、学習コストが低く容易に利用できるようにした軽量版と位置づけられる。ChaKi.NET の開発により、言語データの利用が促進されることを期待したい。なお、ChaKi.NET は開発途上であり、本論文で紹介した点以外にも今後様々な改良を施す予定である。一定の機能を追加後、ChaKi.NET lite は Github 上にて公開予定である。

謝 辞

本研究は国立国語研究所基幹型プロジェクト「実証的な理論・対照言語学の推進」サブプ

プロジェクト「アノテーションデータを用いた実証的計算心理言語学」および、JSPS 科研費 19K13180 の助成を受けたものです。

文 献

- Yuji Matsumoto, Masayuki Asahara, Kiyota Hashimoto, Yukio Tono, Akira Ohtani, and Toshio Morita (2006). “An Annotated Corpus Management Tool: ChaKi.” *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06)*. Genoa, Italy: European Language Resources Association (ELRA).
- Anders Björkelund, Agnieszka Falenska, Xiang Yu, and Jonas Kuhn (2017). “IMS at the CoNLL 2017 UD Shared Task: CRFs and Perceptrons Meet Neural Networks.” *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pp. 40–51. Vancouver, Canada: Association for Computational Linguistics.
- 浅原正幸・金山博・宮尾祐介・田中貴秋・大村舞・村脇有吾・松本裕治 (2019). 「Universal Dependencies 日本語コーパス」 *自然言語処理*, 26:1, pp. 3–36.

関連 URL

コーパス検索アプリケーション 『中納言』	https://chunagon.ninjal.ac.jp/
British National Corpus (BNC)	https://www.english-corpora.org/bnc/
Corpus of Contemporary American English (COCA)	https://www.english-corpora.org/coca/
ChaKi.NET	https://ja.osdn.net/projects/chaki/
ChaKi.NET lite (後日公開予定)	https://github.com/chakidev/chakinet-lite
Grew-match	http://match.grew.fr/
INESS	http://clarino.uib.no/iness
Kontext	http://lindat.mff.cuni.cz/services/kontext/corpora/corplist
PML Tree Query	http://lindat.mff.cuni.cz/services/pmltq/
SETS treebank search	http://depsearch-depsearch.rahtiapp.fi/ds_demo/
Universal Dependencies	https://universaldependencies.org/

アタヤル語群における「言う」の再建

落合 いずみ (帯広畜産大学) †

A reconstruction of “say” in Atayalic languages

Izumi OCHIAI (Obihiro University of Agriculture and Veterinary Medicine)

要旨

アタヤル語 (オーストロネシア語族アタヤル語群) において「要求する」を表す語 *masina* について、その起源をセデック語 (アタヤル語群) との比較から考察し、アタヤル語群祖語を **asa* と再建し、本来は「言う・要求する」を表す語であったことを主張する。アタヤル語の *masina* は 2 通りの形態分析ができ、*m-asina* または *mə-sina* であるが、いずれも語根と推定される ***asina* または ***sina* は、セデック語の同源語が見つからない。一方セデック祖語に再建される「言う」は **asa* であり、同時に「要求する」という意味でもある。アタヤル語の特徴として最終音節のオンセット直後に挿入される特殊な接尾辞 (化石中央接尾辞と呼ばれる) があり、形式は <*in*> である。アタヤル語では祖形 **asa* に <*in*> が挿入され、*as<in>a* が派生され、さらに「要求する」の意味のみに特化したのではないか。だとすればアタヤル語の「言う」を表す形式 *kayal* の由来が問題になる。これはアタヤル語の *kai* 「言葉」(アタヤル語群祖語 **kari*) に化石接尾辞 *-al* が付加して、*kai-al* から *kayal* となったのだろう。アタヤル語群祖語 **asa* にはブヌン語 *asa* 「好む、必要とする、望む」やサイシャット語 *oSa* 「言う」などの同源形式もあり、これらからオーストロネシア祖語は **aSa* と再建されうる。

1. はじめに

アタヤル語群に属するのはアタヤル語とセデック語の 2 つの言語である。アタヤル語群はオーストロネシア語族に属する。Blust (1999) によると、オーストロネシア祖語は 10 に分岐するとされる。アタヤル語群はその 10 分岐の中の 1 つである。他には、東台湾諸語群 (アミ語、カバラン語、シラヤ語などを含む)、西平原諸語群 (ホアニャ語、パポラ語、サオ語などを含む)、北西台湾諸語群 (サイシャット語、バゼツヘ語を含む)、ブヌン語、ツォウ語群 (ツォウ語、サアロア語、カナカナブ語を含む)、ルカイ語、プユマ語、パイワン語があり、アタヤル語群を含めここまでの 9 分岐が台湾本土で話される言語である。残りの 1 分岐がマラヨ・ポリネシア語群であり、台湾本土以外の太平洋、インド洋をまたがる地域に分布する。

小川・浅井 (1935: 21, 599) によると、アタヤル語群に属するアタヤル語とセデック語はそれぞれ、2 つの方言に分けられる。アタヤル語はスコレック方言とツォレ方言に分かれる。

† i.ochiai@obihiro.ac.jp

セデック語はパラソ方言とトゥルク方言に分かれる。本稿においてアタヤル語のデータは先行研究を参照した。セデック語のデータはセデック語パラソ方言については筆者の調査によるもの（または筆者による作例）であり、セデック語トゥルク方言は先行研究を参照した。

本稿は、アタヤル語群において「言う」を表す形式を *əsa と再建するのが目的である。以下2節ではまずセデック語において「言う・要求する」を表す語である *mesa* について検討する。アタヤル語において「要求する」と「言う」はそれぞれ別の形式を用いる。3節ではアタヤル語において「要求する」を表す形式 *məsina* を導入する。4節では、まず化石接辞について説明し、その後でセデック祖語の *m-əsa とアタヤル語の *məsina* が同源であることを示す。5節ではアタヤル語の「言う」を表す形式である *kayal* の由来について検討する。6節ではアタヤル語群祖語に再建される「言う」を表す語根 *əsa について、他の台湾オーストロネシア諸語からサイシャット語とブヌン語の同源形式を挙げ、オーストロネシア祖語に *əSa を再建する。

2. セデック語の「言う」と「要求する」¹

まずセデック語パラソ方言の状況について述べる。「言う」を表す語は同時に「要求する」をも表す。セデック語パラソ方言において「言う・要求する」を表す語は *mesa* である²。これは *m-esa* と分析され、*m-* は静態動詞または動作主態を表す接頭辞と考えられる³。この語 *m-esa* は現実相を示す形式である。これに対し、アタヤル語群において非現実相の形式は命令形と否定辞後の形式を指すが、非現実相の形式では「言う」を表すか「要求する」を

¹ 筆者のフィールド調査によると、セデック語パラソ方言の音素は母音/a e i o u/、二重母音/uy/、子音/p b t d ts k g q s x h m n ŋ l r w y/である。月田 (2009: 56-62) によると、セデック語トゥルク方言は母音/a i u ə/、二重母音/aw ay uy/を持ち、子音ではパラソ方言と同じだが /ts/ がない。ちなみに、パラソ方言の母音 *e* は二重母音 *aw* に遡る。パラソ方言の母音 *e* は次末音節の場合、ə または二重母音 *ay* に遡る。最終音節の場合は *ay* に遡る。

² ちなみに、「言う」を表す *mesa* が文法化を経て伝聞を表す終助詞 *sa* になったことが、落合 (2015: 57) に述べられている。

また、セデック語パラソ方言には、「言う」を表すもうひとつの語があり、形式は *reŋo* (<セデック祖語 *rəŋaw または *rəŋag) である。これは *r<um>ŋo kari* (<AV> speak word) 「言葉を話す」のように用い、「話」「言葉」「考え」といった目的語をとる。そのため「(～を) 話す」がより適切な訳出だろう。一方 *mesa* は *iya so kiya, mesa =ku* (PROH such that <AV> say =1SG.NOM) 「『そのようにしてはいけません』と私は言った。」などのように伝聞の内容や、発言の内容を導入する場合に用いる。

³ ただし、静態動詞の接頭辞 *m-* (正確に言えば *mV-*) が、母音から始まる語根 *esa* に付いているとするなら、音韻規則からすれば、*me-esa* となることが期待される。例えば、語根 *icu* 「恐れる」は *mi-icu* となるように、接頭辞内の母音は、後続の母音と同じものが現れる (落合 2016: 116)。この形式の非現実相は *ki-icu* である。しかし *esa* の場合、期待に反して *m-esa* という形式である。同質の母音の連続 *ee* から同音脱落によって1つの母音が失われたように見える。*k-esa* の場合も同様に *ke-esa* が期待されるが、同音消失が起き *k-esa* になったように見える。

表すかによって接辞が異なる。「言う」を表す場合は、静態動詞に特有の接頭辞である *k-* を付加し *k-esa* となる⁴。「要求する」を表す場合は、動作主態に特有の接頭辞であるゼロ形式を用い、*esa* となる。非現実相に限って言えば、「言う」の場合は静態動詞、「要求する」の場合は動作主態の形態をとることになるが、いずれにせよ語根は *esa* であり、語根の原義は「言う・要求する」である⁵。パラソ方言において同一語根 *esa* が「言う」を意味するか「要求する」を意味するかによって、非現実相が異なる形態を示すことを表1にまとめる。

	現実相	非現実相	語根
言う	<i>m-esa</i>	<i>k-esa</i>	<i>esa</i>
要求する	<i>m-esa</i>	<i>esa</i>	<i>esa</i>

表1 セデック語パラソ方言における「言う」と「要求する」の形態

以下に *m-esa* を「言う」の意味で使った例 (1)、「要求する」の意味で使った例 (2)、「言う」の意味を否定した非現実相の例 (3)、「要求する」の意味を否定した非現実相の例 (4) を挙げる。この他に「要求する」から派生した意味として、(5) に挙げたように「発情する」も見られるが、この場合 (6) に挙げたように「要求する」と同様、非現実相では動作主態の形式をとる⁶。

- (1) *m-esa ka seediq*
SV-say TOP person
「あの人は言う。」
- (2) *m-esa pila ka heya*
AV-demand money TOP 3SG
「彼(彼女)は金を求める。」
- (3) *ini k-esa ka seediq*
neg SV.IRR-say TOP person
「あの人は言わない。言わなかった。」⁷

⁴ 動詞に付加する接辞の形態については落合 (2016: 85, 94–95, 120–121) を参照した。

⁵ このことからセデック語において、何かを口に出して言うことは何かを求めることと表裏一体であることが分かるが、本稿では言うことすなわち要求することという観念の文化的・社会的側面には立ち入らない。

⁶ 略号一覧は以下の通りである。AV: 動作主態、IRR: 非現実相、NEG: 否定辞、NOM: 主格、PROH: 禁止、SG: 単数、SV: 静態動詞、TOP: 主題。

⁷ 例文 (3)、(4)、(6) では否定辞 *ini* が導入されているが、これらの否定文の訳出では現在時制と過去時制の両方の訳が挙げられている。これは、落合 (2016: 53–54) によれば、否定辞後の動詞の形態は非現実相を必要とし、非現実相では時制の区別がないため、現在の意味も過去の意味も表すためである。

- (4) *ini esa pila ka heya*
 NEG AV.IRR-demand money TOP 3SG
 「彼（彼女）は金を求めない。求めなかった。」
- (5) *m-esa ka huling ga*
 AV-demand TOP dog that
 「その犬は発情している。」
- (6) *ini esa ka huling ga*
 NEG AV.IRR-demand TOP dog that
 「その犬は発情していない。いなかった。」

次に、セデック語トゥルク方言における同源語を探り、セデック祖語を再建する。Rakaw 他 (2006: 469) によれば、トゥルク方言において「言う」を表す形式は *māsa* である。この派生形として *mākakasa* 「多く（の人）がそのように言う」が挙げられているが、これは *mākā-k-asa* と分析できるだろう。*mākā-* は相互を表す接頭辞として用いられる *mā-Cā-* (C は語根の初頭子音の重複) に相当するかもしれない⁸。その後ろの *k-* は、パラソ方言において (3) の例文に見られるのと同様に、静態動詞・非現実相を表す *k-* だろう。そして語根は *asa* である。トゥルク方言の「要求する」について、原住民族語言研究發展基金會 (2020) では *meysa* という形式が挙げられている。これは明らかに、「言う」を表す語根 *asa* と関連している。*me-ysa* と分析されるだろう⁹。語根は *ysa* であるが語頭の *y* は *i* が、母音 *a* の直後で半母音に変化したものであると考えられるので、*isa* となる。トゥルク方言では、「要求する」を表す場合、本来 *asa* であった形式を、「言う」との形式的区別を図るために *isa* に変えたのだろうか。表 2 にトゥルク方言における「言う」と「要求する」の形態をまとめる。

	現実相	非現実相 ¹⁰	語根
言う	<i>m-asa</i>	<i>k-asa</i> (?)	<i>asa</i>
要求する	<i>me-ysa</i>	<i>isa</i> (?)	<i>isa</i>

表 2 セデック語トゥルク方言における「言う」と「要求する」の形態

パラソ方言の語根 *esa* 「言う・要求する」とトゥルク方言の語根 *asa* 「言う」の方の形式を

⁸ トゥルク方言の相互形の形態については月田 (2009: 670) を参照した。

⁹ 月田 (2009: 62–63) によると、トゥルク方言において二重母音 *ay* が次末音節にある場合、音声的に [ey] と発音される。そのため、この *meysa* は本来 *maysa* であったが、音声的に [meysa] と聞こえると考えられる。その場合、接頭辞は *ma-* となる。これは静態動詞の接頭辞としてセデック祖語またはそれより早い段階に再建される形式である。トゥルク方言において早期に **ma-asa* という形式が作られ、それが **ma-isa* から *ma-ysa* [meysa] に変わったのだろうか。

¹⁰ トゥルク方言の非現実相の形式は管見の限り得られなかったため、パラソ方言のパターンの類推によって作成した形式を表 2 に挙げた。

比較して再建されるセデック祖語の語根は *asa となる。なぜなら、Ochiai (2018) によるとセデック語パラソ方言の次末音節において通時的に *ə に遡る音素は、e に変化したからである。表 3 にセデック祖語における語根の再建についてまとめる。

パラソ方言	トゥルク方言	セデック祖語
<i>esa</i>	<i>asa</i>	*asa

表 3 セデック祖語における「言う・要求する」の語根の再建

3. アタヤル語の「要求する」¹¹

アタヤル語において「言う」と「要求する」を表すのには、異なる語を用いる。本節と次節では「要求する」を見ていく。小川 (1931: 396) によると、アタヤル語スコレック方言において「要求する」を表す語は *mə-sina* である。この分析では *mə* が接頭辞、*sina* が語根と読み取れる。原住民族語言研究發展基金會 (2020) にはアタヤル語ツオレ方言の下位方言である萬大方言において、*masina* 「要求する」という形式が見られた。小川 (1931: 396) に倣えば *ma-sina* と分析され、語根は同じく *sina* になるだろう。Huang (2018: 273) や落合 (2020: 146) によれば、ツオレ方言の一部では次末よりも前の音節の ə が a になる傾向があり、そのため接頭辞と思しき *ma-* の母音が a になっている。そのため、仮にアタヤル祖語を再建するとすれば *mə-sina になる (次節で述べるが、これは通時的には *m-əs<in>a と分析される)。

これまでの分析を踏まえれば、アタヤル語において「要求する」を表す語根と見なされてきた *sina* はセデック祖語において「言う・要求する」を表す *asa とは異なる語ということになる。そこで探してみたいのが、セデック祖語の形式 *asa との同源である可能性と、それに関わる化石接辞の付加の可能性である。

4. 化石接中辞の付加

アタヤル語群に特徴的な形態変化として化石接辞の付加がある。化石接辞とはその機能が不明であるが、形式的な統一性ならびに付加する位置の統一性が見られる一群の接辞のことである。落合 (2022a: 1-6) はアタヤル語群に見られる化石接辞を化石前方接中辞、化石中央接中辞、化石後方接尾辞の 3 つに分類した。化石前方接中辞はオーストロネシア諸語に広く見られる。一方、化石中央接尾辞はアタヤル語のみに見られ、セデック語には見られない。化石後方接中辞はアタヤル語とセデック語の両方に見られる。さらにアタヤル語群には、化石接尾辞も存在する (小川・浅井 1935: 25、Li 1985: 259、落合 2020)。化石接尾辞については、5 節におけるアタヤル語の「言う」の分析の際に再登場する。

落合 (2022b: 90) によれば、アタヤル語における化石中央接中辞の形式は決まっている。

¹¹ 筆者のフィールド調査 (2018 年から 2019 年にかけて) によると、アタヤル語スコレック方言の音素は母音 /a e i o u ə/, 二重母音 /aw ay uy/, 子音 /p β t k ɣ q ʔ s x h z r l m n ŋ y w/ であった。母音 o と e はそれぞれ二重母音 aw と ay に遡ることがこれまでの調査において観察できた。Huang (1995: 16-17) におけるアタヤル語ツオレ方言の音素目録によると、ツオレ方言では /ʁ/ を持ち、/ə/ を持たない。

一音節の V(C)の構造を持ち、母音は *i* であり、後続の子音は *l* または *n* で現れるため、<*i*>、<*il*> または <*in*> で現れる。挿入位置は最終音節のオンセットの直後である。例えば、語根が CVCVC の音節構造をもっていたとして、化石中央接中辞の <*il*> が挿入されると CVC<*il*>VC になる。このようにして作られた語の例として、「サトウキビ」が挙げられる。「サトウキビ」はセデック語では *sibus* というが、アタヤル語では *bilus* である。アタヤル語では早い時期に *sabus* だったと考えられるが (オーストロネシア祖語は *CəbuS)¹²、化石中央接中辞<*il*>の挿入によって *səb<il>us* になり、アクセントがおかれる次末音節より前の音声的に弱い前次末音節が脱落して *bilus* になったと述べられている。

「要求する」に話を戻す。アタヤル語スコレック方言の *məsina* はこれまで *mə-* と *sina* に分節されてきたが、*m-* と *əsina* に分節される可能性があるのではないか。だとすれば、接頭辞は *m-* となる。そして語根に相当する部分は *əsina* になるが、これはセデック祖語と同一の形式である *asa* から化石中央接中辞の付加により派生されたのではないか。もし *əsina* が語基だったとすれば、*asa* に対して化石中央接中辞の <*in*> が、最終音節 *sa* のオンセットである *s* の直後に挿入されて、*əs<in>a* となっていることになる。この語基、*əs<in>a* に対し接頭辞の *m-* が動作主態を表す接頭辞として付加され、*m-əs<in>a* となったと考えられる。この語基をアタヤル祖語に再建するなら **əs<in>a* になる。この再建形と、セデック祖語の **əsa* から再建されるアタヤル語群祖語の「言う・要求する」は **əsa* である。アタヤル語群祖語の形式の再建に関わるデータを表 4 にまとめる。

セデック祖語	アタヤル祖語	アタヤル語群祖語
* <i>əsa</i>	* <i>əs<in>a</i>	* <i>əsa</i>

表 4 アタヤル祖語における「言う・要求する」の語根の再建

ここまでアタヤル語の「要求する」について通時的な分析をしてきたが、共時的な観点から言うと、*məsina* を *mə-* と *sina* に分節し、*sina* を語根ととらえていると考えられるデータが見受けられる。まず、Egerod (1980: 626) のアタヤル語辞典において、見出し語として *sina* が現れる。このことから *sina* が語根と考えられていることがわかる。この派生形として挙げられているのが *məsina* の他に、*sənina* や *sənan* である。前者は *s<ən>ina* と分析される (非動作主態・対象主語・過去形を表す形式)。後者は *sina-an* から、音声的に弱い音節である前次末音節の母音が曖昧母音になり *səna-an* へ変化し、同音消失で *səna-n* になったと考えられる (非動作主態・場所主語を表す形式)。どちらの形式も、*əsina* ではなく *sina* を語根と見なして屈折させた形式である。ただし、一例だけ *isənan* という形式が見られた。これは *isəna-an* が同音消失により、*isəna-n* になったものだろう。語頭母音として *i* が付いているが、これは本来の語基であった *əsina* における語頭母音の *ə* の名残ではないだろうか。そうだとすれば、語頭の *ə* が *i* に変わっている。2 節で述べたが、セデック語トゥルク方言では「言う」ではなく「要求する」の意味の場合に *asa* から *isa* に変化した (表 2 参照)。アタヤル語にも同様の変化が起きた可能性を示す語例である¹³。

¹² オーストロネシア祖語の形式は Blust and Trussel (2010) からの引用である。

¹³ なぜ *ə* から *i* へ変化し、他の母音に変化しなかったかについてであるが、これは *ə* が子音 *s* と隣接していることに関わるのではないか。4 節で述べたようにセデック語の *sibus* (パラン方言とトゥルク方言は同一形式) はオーストロネシア祖語の **Cəbus* の反映形である。

5. アタヤル語の「言う」

前節では、アタヤル語において「要求する」を表す語がセデック語の「言う・要求する」と同源語であることを述べた。本節ではアタヤル語において、「要求する」とは異なる形式を持つ「言う」を表す語の由来について検討する。

小川 (1931: 22) において、アタヤル語スコレック方言の「言う」は $k\langle\text{əm}\rangle\text{ayal}$ であるため、語根は *kayal* であることが分かる。原住民族語言研究發展基金會 (2020) ではアタヤル語ツオレ方言の下位方言の1つである汝水下位方言における「言う」の語根は *kaal* である。Li (1981: 265) によれば、汝水下位方言では同質の母音間において *r に遡る子音 *y* が脱落する傾向が見られるため、*kayal* がより古い形式であり、母音 *a* に挟まれた *y* が脱落して *kaal* となったのだろう。アタヤル祖語を再建するとすれば **kayal* になる。

この形式には、同音異義語がある。スコレック方言においてもツオレ方言においても *kayal* (小川 1931: 209) または *kaal* (原住民族語言研究發展基金會 2020) は「空」をも意味する。しかも、「空」の方はセデック語において同源形式が見られる。セデック語パラン方言では *karac*、セデック語トゥルク方言では *karat* (Rakaw 他 2006: 351) である¹⁴。

では、アタヤル語において「言う」を表す *kayal* はどこから来たのだろうか。そこで考えてみたいのが、またも化石接辞の付加である。今回は化石接尾辞が付加した可能性を検討する。アタヤル語群に特徴的に見られる化石接尾辞にはいくつかの形式が知られている。

例えば *-iq* が挙げられるが、これには *-niq* (Li 1985: 259)、*-liq* (落合 2020) などの変化形も見られる。このほか落合 (2022c) によると *-ur* という化石接尾辞があるが、これには *-hur* という語頭子音を伴った変化形も見られる。これらは一音節から成り、-(C)VC の型を持つ。語根が母音で終わる場合は *-CVC* が付加し、語根が子音で終わる場合には *-VC* が付加すると考えられる。

アタヤル語スコレック方言に *kai* という語があり、意味は「言葉」である (小川 1931: 138)。これは Li (1981: 287) によると、アタヤル語群祖語 **kari* の反映形である¹⁵。これはオーストロネシア祖語 **kaRi* の反映形である¹⁶。原住民族語言研究發展基金會 (2020) によるとアタヤル語ツオレ方言の下位方言の1つである萬大下位方言では *ke* である。より古い形式の *kai* における母音連続 *ai* が *e* に変化している。ちなみにセデック語ではパラン方言も *kari*、トゥルク方言も *kari* (Rakaw 他 2006: 352) である。このアタヤル語の *kai* に対し、化石接尾辞の *-al* を付加したのではないか。

この傍証としてサイシャット語の「言う」が挙げられる。Zeitoun et al. (2015: 548) では、サイシャット語において「言葉」を表す語である *kai'* [kai?] が挙げられるが、もちろん Blust

次末音節の母音は *a* が期待されるのだが、なぜか *i* に変わった。しかも子音 *s* が直前にある。アタヤル語群において子音 *s* と母音 *i* の間に親和性があると見なせるのではないか。

¹⁴ これら形式から再建されうるアタヤル語群祖語の「空」は **karad* である。Li (1981: 254) によれば、アタヤル語群祖語 **d* は、アタヤル語では語末の位置で脱落ということである。その点では、アタヤル語の *kayal* は語末が *l* で反映されるので例外的であるが、**d* から *l* へ変化は起こりそうな変化である。

¹⁵ Li (1981: 287) の再建形では語末子音として *ʔ* が添えられているが (**kariʔ*)、本稿では母音終わりの語に対し、音声的に生じた子音であり音素的ではないと見なしている。

¹⁶ オーストロネシア祖語の形式は Blust and Trussel (2010) を参照した。

and Trussel (2010) によると、オーストロネシア祖語 *kaRi の反映形である。Zeitoun et al. (2015: 548) はこの語根に対し、接頭辞の *ma'ya-* が付加し、*ma'ya-kai'* となることで「話す」という動詞が作られるとする。つまり、サイシヤット語では「言葉」を表す語から「話す」が派生されるのであり、しかもサイシヤット語で「言葉」を表す語 *kai'* は、アタヤル語の *kai* と同源の形式である。ちなみに、6 節では、サイシヤット語におけるもう 1 つの「言う」の形式も登場する。

次に、Li (1985: 259) にはアタヤル語における化石接辞として *-al* が見られることが述べられている。例えば Li (1985: 259) によるとツオレ方言汶水低位方言では、「見る」を表す本来の形式 *kita* に対し *-al* を付加して *kita-al* とする¹⁷。また、Ochiai (2019) によると、*atayal* 「アタヤル族」は、*ita* (一人称複数包括形) に対し、*-yal* という *-al* と同類の化石接辞が付いた形式である (*ita* > *ita-yal* > *ata-yal* > *ata-yal*)。

アタヤル語において「言う」を表す *kayal* は、*kai* に対し *-al* が付き *kai-al* となり¹⁸、再音節化が起こって *kay-al* となったのではないか。そしてこれを「言う」の意味で用いたのではないか。「言葉」と「言う」の間に意味的関連が見られることは言うまでもない。アタヤル語では本来、セデック祖語と同様の **əsa* という形式を「言う」として用いていたと予想され、それが後に **kay-al* に置き換わったのだろう。古い形式である **əsa* は、アタヤル語において「言う」の意味では他の語に取り換えられ、「要求する」では化石中央接中辞が挿入されて **əs<in>a* になった。

「言う」と「要求する」について、表 5 にセデック祖語とアタヤル祖語の語根 (またはアタヤル語群については化石接辞の付加により派生された語基) の形式とそれらから推定されるアタヤル語群祖語の語根形式を示す。

	セデック祖語	アタヤル祖語	アタヤル語群祖語
言う	* <i>əsa</i>	(* <i>əsa</i> >) * <i>kay-al</i>	* <i>əsa</i>
要求する	* <i>əsa</i>	* <i>əs<in>a</i>	* <i>əsa</i>

表 5 アタヤル語群祖語における「言う・要求する」の再建

¹⁷ ただし、ツオレ方言汶水低位方言では同質母音間の **r* に遡る子音が脱落する変化がある。そのため *-al* ではなく、より古い形式で言うならば **-ral* という子音 **r* から始まる化石接尾辞が付加した可能性がある。その場合、化石接辞を付加すると *kita-ral* となり、子音 *r* が脱落したことになる。

¹⁸ 母音終りの語には子音始まりの化石接尾辞 *-Cal* が付くはずであるが、ここでは母音始まりの *-al* が付いているようである。これは、語根 *kai* における母音連続の後半母音の *i* が、化石接尾辞付加の時点で半母音 (子音として扱われると考えられる) と捉えられていたからだろう。または、*kai* に対し、*-yal* という化石化石接尾辞が付き、*kai-yal* になり、再音節化で *kay-yal* になり同音消失で *kay-al* になった可能性もある。脚注 17 に関連するが、化石接尾辞が、この *-yal* はより古い形式として **-ral* であった可能性が高い。Li (1981: 265) によると、スコレック方言では **r* が *y* に変わる。そのため化石接尾辞が **-ral* であったならスコレック方言で *-yal* に変化したと推察される。

6. オーストロネシア祖語への再建

アタヤル語群祖語に再建された *əsa「言う・要求する」は、他のオーストロネシア諸語において同源語が見られるのだろうか。まず、葉 (2000: 137) によるとアタヤル語と隣接するサイシャット語に *k<om>oSa'*「言う」という形式があり、語根は *koSa'* [koʃaʔ] であることが分かる。これは、語頭子音の *k* を除いた *oSa'* [oʃaʔ] の部分が、アタヤル語群祖語の *əsa に相当するように見える。Zeitoun et al. (2015: 270) によると、サイシャット語において静態動詞を表す接頭辞として *k-* が挙げられている。そのため *koSa'* の語頭子音 *k* は、通時的に考えれば、静態動詞を表す *k-* に由来するだろう。サイシャット語におけるこの接頭辞は表 1 におけるセデック語パラソ方言の「言う」の非現実相 *k-esa* における静態動詞接頭辞 *k-* に相当するものである。

次に、Nihira (1988: 16) には、セデック語に隣接する言語であるブヌン語において、*asa* という形式が見られる。意味は「好む、必要とする、望む」などが挙げられているが、これらの意味は「要求する」に繋がる。これも、アタヤル語群祖語 *əsa「言う・要求する」の同源語と考えられる。

アタヤル語群祖語の *əsa、サイシャット語の *koSa'* の歴史的語根 *oSa'*、ブヌン語の *asa* が同源語だと考えられる。これらからオーストロネシア祖語を再建するなら、語中子音は *S となる。この子音は各言語において規則的な反映を示している (アタヤル語群祖語 *s、サイシャット語 *S* [ʃ]、ブヌン語 *s*)¹⁹。問題になるのは語頭母音である。アタヤル語群祖語では *ə である。このアタヤル語群祖語の母音は Li (1981: 275) ではオーストロネシア祖語の *ə に遡るとする。サイシャット語では *o* である。このサイシャット語の母音は Li (1978: 140) ではオーストロネシア祖語 *u に遡るとする。ブヌン語では *a* である。このブヌン語の母音は Li (1988: 493) によるとオーストロネシア祖語の *a に遡るとする²⁰。

語頭母音について、アタヤル語群祖語の語形は *ə、ブヌン語の語形は *a、サイシャット語の語形は *u に遡ることを示唆し、一致を見ない。種々の母音に変化する可能性が高い母音はどれかを考えるなら、*ə を暫定的に建てておくのが無難に思われる。だとすれば、ブヌン語において *ə がなぜか *a* に、サイシャット語において *ə がなぜか *o* に変わったことになる。

アタヤル語群祖語における「言う・要求する」、サイシャット語における「言う」、ブヌン語における「好む、必要とする、望む」の同源形式と、それらから再建されるオーストロネシア祖語の形式を表 6 にまとめる。

¹⁹ オーストロネシア祖語に再建される語中子音は *s ではなくて *S であるのは、アタヤル語群が *s であることが決め手になる。もし、オーストロネシア祖語で *s なら、アタヤル語群祖では *h になることが期待されるからである。しかも、Blust and Trussel (2010) によればオーストロネシア祖語に *əsa「1」という語が存在する。本稿で再建した *əSa「言う・依頼する」とは最小対の関係にある。

²⁰ Li (1988: 493) によるとオーストロネシア祖語の *ə がブヌン語において *a* で反映されることもあるが、その場合は子音 *q* に隣接しているという条件が付いている。*asa* の場合は後続の子音は *s* であるのでこの条件に当てはまらない。

アタヤル語群祖語 「言う・要求する」	サイシヤット語 「言う」	ブヌン語 「好む、必要とする、望む」	オーストロネシア祖語 「言う・要求する」
*əsa	oSa' [oʃaʔ] ²¹	asa	*əSa

表 6 オーストロネシア祖語における「言う・要求する」の再建

7. おわりに

本稿はアタヤル語群祖語において「言う」を表す語を再建した。結果として「言う」と「要求する」は表裏一体であり、アタヤル語群祖語には *əsa という形式が「言う・要求する」の意味で再建された。セデック語においてこの形式・意味は保存されている。一方アタヤル語において *əsa はそのままの形式で保存されておらず、「言う」は kai「言葉」を基に化石接尾辞 *-al* の付加によって造られた形式 *kay-al* を用いる。アタヤル語の「要求する」の場合は、化石接中央接中辞 <in> を挿入した形式 *əs<in>a に変わった。

アタヤル語群祖語の *əsa は、サイシヤット語とブヌン語に同源語が見られた。サイシヤット語では「言う」の意味で *k-oSa'* という形式が用いられ、ブヌン語では「要求する」の意味で *asa* という形式が用いられる。これら 3 つの言語 (群) の形式・意味の比較を基に、オーストロネシア祖語において *əSa「言う・要求する」が再建された。

謝 辞

本稿に対し野島本泰氏からご助言をいただいたことに感謝する。しかし本稿の不備はすべて筆者の責任である。

文 献

- Blust, Robert (1999) Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In: Elizabeth Zeitoun and Paul Jen-kuei Li (eds.) *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, 31–94. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Blust, Robert and Stephen Trussel (2010) *Austronesian Comparative Dictionary, Web Edition*. <http://www.trussel2.com/ACD/> [2022 年 8 月アクセス].
- Egerod, Søren (1980) *Atayal-English dictionary, vol. 1–2*. London: Curzon.
- 原住民族語言研究發展基金會 (2020) 原住民族語 E 樂園 <http://web.klokah.tw> [2022 年 8 月アクセス].
- Huang, Hui-chuan J. (2018) The nature of pretonic weak vowels in Sqliq Atayal. *Oceanic Linguistics* 57(2): 265–288.
- Huang, Lillian M. (1995) *A Study of Mayrinax Syntax*. Taipei: Crane.
- Li, Paul Jen-kuei (1978) A comparative vocabulary of Saisiyat dialects. *Bulletin of the Institute of*

²¹ サイシヤット語に見られる語末子音の声門閉鎖音は、母音終りの語に対し、発話時に音声的に加えられたものと考えた。

- History and Philology, Academia Sinica* 49: 133–199.
- Li, Paul Jen-kuei (1981) Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 52(2): 235–301.
- Li, Paul Jen-kuei (1985) The position of Atayal in the Austronesian family. In: Andrew Pawley and Lois Carrington (eds.) *Austronesian linguistics at the 15th pacific science congress*, 257–280. Canberra: Pacific Linguistics.
- Li, Paul Jen-kuei (1988) A comparative study of Bunun dialects. *Bulletin of the Institute of History and Philology, Academia Sinica*, 59(2): 479–508.
- Nihira, Yoshiro (1988) *A Bunun vocabulary: A language of Formosa*. Tokyo: Ado-in Kabushiki Kaisha. Third edition.
- 小川尚義 (1931) 『アタヤル語集』 台北: 台湾総督府.
- 小川尚義・浅井恵倫 (1935) 『原語による台湾高砂族伝説集』 台北: 台北帝国大学言語学研究室.
- 落合いずみ (2015) 「セデック語パラン方言の談話資料『ソメコ』と終助詞の用法」『地球研言語記述論集』 7: 39–64.
- 落合いずみ (2016) 「セデック語パラン方言の文法記述と非意志性接頭辞の比較言語学的研究」博士論文、京都大学.
- Ochiai, Izumi (2018) Historical reduplication in Seediq. *Kyoto University Linguistic Research* 37: 23–40.
- Ochiai, Izumi (2019) Atayal: The origin of the tribal name. 『国立清華大學世界南島暨原住民族中心電子報』 1: 52–54.
- 落合いずみ (2020) 「アタヤル語群における「肩」の再建」『アジア・アフリカ言語文化研究』 100: 141–153.
- 落合いずみ (2022a) 「セデック語の方言比較から浮き彫りになる化石接中辞」『アイヌ・先住民研究』 2: 1–29.
- 落合いずみ (2022b) 「アタヤル語の『サトウキビ』に起きた特異な形態変化」『北方人文研究』 15: 85–97.
- 落合いずみ (2022c) 「パゼッへ語とアタヤル語群の『年上キョウダイ』の再建」『北海道方言研究会会報』 98: 27–34.
- Rakaw, Lowsi, Jiru Haruq, Yudaw Dangaw, Yuki Lowsing, Tudaw Pisaw, and Iyuq Ciyang 編 (2006) 『太魯閣族語簡易字典』 秀林郷: 秀林郷公所.
- 月田尚美 (2009) 「セデック語（台湾）の文法」博士論文、東京大学.
- 葉美利 (2000) 『賽夏語参考語法』 台北: 遠流.
- Zeitoun, Elizabeth, Tai-hwa Chu, and Lalo a tahesh kaybaybaw (2015) *A study of Saisiyat morphology*. Honolulu: University of Hawai‘i.

編集後記

国立国語研究所第4期基幹型プロジェクト「実証的な理論・対照言語学の推進」(2022-2027年度)のプロジェクトリーダーの浅原と申します。本プロジェクトは研究系理論・対照グループのサブプロジェクトの取りまとめを行うほか、二つを取組を行います。

一つ目は理論・対照グループのイベントの開催です。国語研第3期においては、理論・対照研究領域による『Prosody and Grammar Festa』を神戸大学と共同で開催しておりました。同イベントを引き継いで、本イベント『Evidence-based Linguistics Workshop』を企画しました。新体制においてどのようなイベントを開催するかを検討し、研究所内の共同研究プロジェクトに参画されている共同研究員のみならず、どなたでも発表申込が可能なものとししました。また、ハイブリッド開催のために必要な準備を進めました。

二つ目は言語学のオープンサイエンス化です。近年、内閣府主導のもと、国際動向を踏まえたオープンサイエンス推進のための方策が検討され、様々な国内施策が進められています。言語学の分野では、紙の論文や書籍の出版が業績として認められた時代が長く、従来の業績を前提とした学術情報流通システムが強固であるために、研究成果物の共有化が進んでいない状況にあります。言語系の学会においては、制約の厳しい規定・著作権譲渡事項を設定していることにより、プレプリントサーバ・学位論文なども含めて二次投稿に対する制限を課しています。本ワークショップでも、17件の発表申込のうち発表論文の提出が2件のみでした。本ワークショップの発表論文集は、著作者に不利益のないように「著作権を原著者が保持したまま、Creative Commons ライセンスにより文献を公開する」という施策をとり、著作権譲渡による二次投稿の制限を回避します。言語系学会の各規定に対する問題提起を行うとともに、働きかけを進めていきたいと思っております。

浅原正幸 (国立国語研究所)