

「日本の危機言語語彙 データベース」

セリック・ケナン（日本学術振興会特別研究員）

籠宮 隆之（国立国語研究所）

宮川 創（国立国語研究所）

木部 暢子（人間文化研究機構）

12月4日（日）・国立国語研究所

0. 本発表の目的

- 「日本の危機言語語彙データベース」を紹介すること

0. 役割分担

「日本の危機言語語彙データベース」の改修事業における分担

- 発注監修：木部暢子（2022年3月）
- 発注関係：籠宮、セリック
- データ設計：セリック、籠宮
- データ編集：セリック
- サーバ管理：籠宮（2022年3月まで）、宮川（2022年4月より）
- 本発表：セリック

1. これまでの電子語彙データの試み

	名称	公開	継続	レコード数 (推測)	文献
A	琉球語音声データベース	2005以前	終了	40,000	狩俣2005
B	宮古語電子辞書	2016年3月	終了	10,000	
C	日本の消滅危機言語DB	2018年5月	終了	10,000	
D	日琉諸語諸方言の電子辞書	2022年3月	継続		カルリノ2022
E	大琉球語辞典	2023年?月	未公開	140,000	

- 海外だと、『Dictionaria』、「Pangloss」など多数

1.これまでの電子語彙データの試み

	名称	公開	継続	レコード数 (推測)	文献
A	琉球語音声データベース	2005以前	終了	40,000	狩俣2005
B	宮古語電子辞書	2016年3月	終了	10,000	
C	日本の消滅危機言語DB	2018年5月	終了	10,000	
D	日琉諸語諸方言の電子辞書	2022年3月	継続		カルリノ2022
E	大琉球語辞典	2023年?月	未公開	140,000	

- Aは先駆的、B、Cはデータ設計的に問題、Dは属性・検索機能を工夫、Eは未詳

2. 語彙データの構築

- 語彙データを収集・整備する目的は？

2. 語彙データの構築

- 語彙データを収集・整備する目的は？
 - ✗ 「辞典」を作るのは目的ではない（紙・電子問わず）！
 - ✓ ある方言の語彙体系を適切に記述することが目的

- 「辞典」は単なるデータの可視化に過ぎない

方言 (仮名) 方言 (国際音声記号) の 詳細 で 調べ替え

方言 (仮名)	方言 (国際音声記号)	意味	品詞	ア型	地点
あじ	ɕi	あじ【名】	名	m1	宮古水納
あじ	ɕi	あじ【接詞】	名	m1	宮古水納
あじでい	ɕi	(組や欄を)じゅうじにかける。(たすきを)かける【十字にかける。(たすき)をかける】	動	c	宮古水納
あじと	ɕi	あじ【接】	名	m1	宮古水納
あじと	ɕi	あじ【接】	名	m1	宮古水納

pari	id	地点	表記1	表記2	IPA
6269	6269				
5287	1	5157	水納	あ	a
1	1	1	水納	あー	a:
2	1	2	水納	あーい	a:i
5001	1	4847	水納	あーう	a:u
5293	1	5173	水納	あーか]い	a:kai
3	1	3	水納	あーく]	a:gu:
5294	1	5174	水納	あーく]る	a:guru
4	1	4	水納	あーさ	a:sa
5295	1	5175	水納	あー]しゃじゅー	a:cadzu:
5	1	5	水納	あー]す	a:sɕ
5865	1	5746	水納	あー]す	a:sɕ
5296	1	5176	水納	あー]すぎん	a:sɕgin
6	1	6	水納	あー]だ	a:da



2. 語彙データの構築

- ここで言う「語彙データ」は
 - 一般的な大規模の琉和辞典に含まれるデータ
 - +関連の音声ファイル（単独、ア資料、例文）とその属性

2. 語彙データの構築

- 適切なデータ設計をするなら、語彙体系を記述するためには、Relational Database (RDB) のモデルが必須 (Cf. 『Dictionaria』の投稿規定)
- 適切なデータ設計とは？ => 「1セル・1情報」の原理
 - 性質の異なる情報 (ATTRIBUTE・属性) は混同しない
 - 同一の属性に対して複数の値がある場合は、同じレコードで管理しない

2. 語彙データの構築

- RDBはKEYで結ばれている複数のデータテーブル
- RDBを設計するときに、次の2点を決定する必要がある
 - 記述単位の定義：何を記述するRDB？
 - 構造化：記述する単位に対してどのような属性を持たせるか？

2. 語彙データの構築

事例（上野2016）

項目	与那国方言	項目	与那国方言
A		着物	'Nnani] C (その(2)の「麻」で「麻の着物」を bu:'Nnami_としたが, bu:'Nnani_の入力ミス)
ほくろ	'ada= A (痣との区別なし)	鍬	PaNngai= A
明かり	'agai= A	子	'agami] C, 'agamiTi_B
東	'aNngai= A	小豆	'agamami= A
陸	dinaga_B (地中=土地, 地面)	つむじ	midaTi= A
蟻	'aja] C	南瓜	naNKu= A
友	duCji= A	北	niCji= A
匂い	kada= A, kadai= A	虹	'aminumja_B (「雨の庭」らしい)
釜	haNgama= A	草履	saba= A
恋人	?	魚取り網	'aN_B, 'ijuTui]aN_C + B
食糧・主食	hamai= A, 'i_B	鶉	'uduru] C
額	Tai= A		
煤	CjiCji] C (その(1)で, 最初 CjiCji], 再調査でCjiCji=		

2. 語彙データの構築

事例（上野2016）

- テーブルは1点
- 用意されている欄は2つ
- 情報の種類は6つ以上
(全体でおろさく9つぐらい)
- 同一のレコードに複数の値

=> 構造化してみると？

表1 松森類別語彙リスト

項目	与那国方言	項目	与那国方言
A		着物	'Nnani] C (その(2)の「麻」で「麻の着物」を bu:'Nnami_としたが, bu:'Nnani_の入力ミス)
ほくろ	'ada= A (痣との区別なし)	鍬	PaNngai= A
明かり	'agai= A	子	'agami] C, 'agamiTi_B
東	'aNngai= A	小豆	'agamami= A
陸	dinaga_B (地中=土地, 地面)	つむじ	midaTi= A
蟻	'aja] C	南瓜	naNKu= A
友	duCji= A	北	niCji= A
匂い	kada= A, kadai= A	虹	'aminumja_B (「雨の庭」らしい)
釜	haNgama= A	草履	saba= A
恋人	?	魚取り網	'aN_B, 'ijuTui]aN_C + B
食糧・主食	hamai= A, 'i_B	鶉	'uduru] C
額	Tai= A		
煤	CjiCji] C (その(1)で, 最初 CjiCji], 再調査でCjiCji=		

2. 語彙データの構築

事例（上野2016）

記述単位の特定

- ・ 方言形（回答形）

項目	与那国方言	項目	与那国方言
A		着物	'Nnani] C (その(2)の「麻」で「麻の着物」を
ほくろ	'ada=A (痣との区別なし)		bu:'Nnami_としたが,
明かり	'agai=A		bu:'Nnani_の入力ミス)
東	'aNngai=A	鍬	PaNngai=A
陸	dinaga_B (地中=土地, 地面)	子	'agami] C, 'agamiTi_B
蟻	'aja] C	小豆	'agamami=A
友	duCji=A	つむじ	midaTi=A
匂い	kada=A, kadai=A	南瓜	naNKu=A
釜	haNgama=A	北	niCji=A
恋人	?	虹	'aminumja_B (「雨の庭」らしい)
食糧・主食	hamai=A, 'i:_B	草履	saba=A
額	Tai=A	魚取り網	'aN_B 'ijuTui'aN_C+B
煤	CjiCji] C (その(1)で, 最初 CjiCji], 再調査でCjiCji=	鶉	'uduru] C

2. 語彙データの構築

事例（上野2016）

属性の特定

- ・ 調査票
- ・ 系列
- ・ 調査項目
- ・ ピッチ付き語形
- ・ アクセント型
- ・ 備考

表1 松森類別語彙リスト

項目	与那国方言	項目	与那国方言
A		着物	'Nnani] C (その(2)の「麻」で「麻の着物」を bu:'Nnami_としたが, bu:'Nnani_の入力ミス)
ほくろ	'ada=A (痣との区別なし)	鍬	PaNngai=A
明かり	'agai=A	子	'agami] C, 'agamiTi_B
東	'aNngai=A	小豆	'agamami=A
陸	dinaga_B (地中=土地, 地面)	つむじ	midaTi=A
蟻	'aja] C	南瓜	naNKu=A
友	duCji=A	北	niCji=A
匂い	kada=A, kadai=A	虹	'aminumja_B (「雨の庭」らしい)
釜	haNgama=A	草履	saba=A
恋人	?	魚取り網	'aN_B 'ijuTui'aN_C+B
食糧・主食	hamai=A, 'i_B	鶉	'uduru] C
額	Tai=A		
煤	CjiCji] C (その(1)で, 最初 CjiCji], 再調査でCjiCji=		

2. 語彙データの構築

事例（上野2016）

構造化済みデータ

調査票	系列	項目	語形	型	備考
松森類別語彙	A	ほくろ	'ada=	A	痣との区別なし
松森類別語彙	A	明かり	'agai=	A	
松森類別語彙	A	東	'aNngai=	A	
松森類別語彙	A	陸	dinaga_	B	地中 = 土地, 地面
...
松森類別語彙	A	魚取り網	'aN_	B	...
松森類別語彙	A	魚取り網	'ijuTui]aN_	C+B	

2. 語彙データの構築

事例（上野2016）

構造化済みデータ（**拡張版**）

調査票	系列	項目	読み	同源	語形	拍数	型	単位	備考
松森類別語彙	A	ほくろ	ほくろ	0	'ada=	2	A	1	痣との区別なし
松森類別語彙	A	明かり	あかり	1	'agai=	3	A	1	
松森類別語彙	A	東	ひがし	1	'aNngai=	4	A	1	
松森類別語彙	A	陸	りく	0	dinaga_	3	B	1	地中 = 土地, 地面
...
松森類別語彙	A	魚取り網	さかなとりあみ	0	'aN_	2	B	1	...
松森類別語彙	A	魚取り網	さかなとりあみ	0	'ijuTui]aN_	6	C+B	2	

2. 語彙データの構築

事例（上野2016）

構造化済みデータ（拡張版）

問題点は異なる調査項目に対して
同一の語形の場合？

調査票	系列	項目	読み	同源	語形	拍数	型	単位	備考
松森類別語彙	A	ほくろ	ほくろ	0	'ada=	2	A	1	痣との区別なし
松森類別語彙	A	明かり	あかり	1	'agai=	3	A	1	
松森類別語彙	A	東	ひがし	1	'aNngai=	4	A	1	
松森類別語彙	A	陸	りく	0	dinaga_	3	B	1	地中 = 土地, 地面
...
〇〇〇	〇	地面	じめん	0	dinaga_	3	B	1	
松森類別語彙	A	魚取り網	さかなとりあみ	0	'aN_	2	B	1	...
松森類別語彙	A	魚取り網	さかなとりあみ	0	'ijuTui]aN_	6	C+B	2	

2. 語彙データの構築

事例（上野2016）

構造化済みデータ（拡張版）

1. 記述単位を「調査項目∩語形」として定義するか
2. 語形のテーブルを導入するか

調査票	系列	項目	読み	同源	語形	拍数	型	単位	備考
松森類別語彙	A	ほくろ	ほくろ	0	'ada=	2	A	1	痣との区別なし
松森類別語彙	A	明かり	あかり	1	'agai=	3	A	1	
松森類別語彙	A	東	ひがし	1	'aNngai=	4	A	1	
松森類別語彙	A	陸	りく	0	dinaga_	3	B	1	地中 = 土地, 地面
...
〇〇〇	〇	地面	じめん	0	dinaga_	3	B	1	
松森類別語彙	A	魚取り網	さかなとりあみ	0	'aN_	2	B	1	...
松森類別語彙	A	魚取り網	さかなとりあみ	0	'ijuTui]aN_	6	C+B	2	

語形ID	音韻形	ピッチ付き形	拍数	型	単位
1	'ada	'ada=	2	A	1
2	'agai	'agai=	3	A	1
3	'aNngai	'aNngai=	4	A	1
4	'dinaga	dinaga_	3	B	1
...
n	'aN	'aN_	2	B	1
n+1	'ijuTuiaN	'ijuTui]aN_	6	C+B	2

1. 記述単位を「調査項目の語形」にするか
2. 語形のテーブルを導入する

②調査項目のテーブル

①語形のテーブル

語形ID	調査票	系列	項目	読み	同源
1	松森類別語彙	A	ほくろ	ほくろ	0
2	松森類別語彙	A	明かり	あかり	1
3	松森類別語彙	A	東	ひがし	1
4	松森類別語彙	A	陸	りく	0
...
4	〇〇〇	〇	地面	じめん	0
n	松森類別語彙	A	魚取り網	さかなとりあみ	0
n+1	松森類別語彙	A	魚取り網	さかなとりあみ	0

一対他

テーブルごとに1つの記述単位

2. 語彙データの構築

- RDBを構築する際
 - 何を記述するか（最上位の記述単位は何か）？
＜＝＞何に対してアノテーションをしたいか？
 - どのような属性があるのか？
＜＝＞研究目的に対して、どのようなアノテーションを施したいか？
 - 個別の属性は何の属性なのか（何にぶら下がるか）？
＜＝＞データの全体構造はどうなっているか？

3. 日琉諸方言の語彙データを構築する

- 日琉諸語の一方言の語彙体系を記述するためのデータ設計は？
 - 満たしたい要件
 - A. 最も複雑な琉和辞典が編集可能（<=データ設計）
 - B. 音声付き（単独、ア資料、例文）（<=データ設計）
 - C. 安易な可視化（電子・紙）（<=ユーザビリティ）
 - D. 人間には安易に編集可能（<=ユーザビリティ）

3. 日琉諸方言の語彙データを構築する

- 日琉諸語の一方言の語彙体系を記述するためのデータ設計は？
 - 満たしたい要件
 - A. 最も複雑な琉和辞典が編集可能（<=データ設計）
 - B. 音声付き（単独、ア資料、例文）（<=データ設計）
 - C. 安易な可視化（電子・紙）（<=ユーザビリティ）
 - D. 人間には安易に編集可能（<=ユーザビリティ）

上限なしの「音声付」

3. 日琉諸方言の語彙データを構築する

- 日琉諸語の一方言の語彙体系を記述するためのデータ設計は？
 - 満たしたい要件
 - A. 最も複雑な琉和辞典が編集可能（<=データ設計）
 - B. 音声付き（単独、ア資料、例文）（<=データ設計）
 - C. 安易な可視化（電子・紙）（<=ユーザビリティ）
 - D. 人間には安易に編集可能（<=ユーザビリティ）

これは極めて重要

3. 日琉諸方言の語彙データを構築する

- 一案は「語彙データフォーマット（仮）」
=>セリック、大浦（2022）『みんなふつ語彙集』に適応
(<http://doi.org/10.15084/00003679>)

当然、

- 細かいデータ設計は方言ごとに調整する必要がある
- 「正解」ではなく、（うまく行った？）一事例に過ぎない

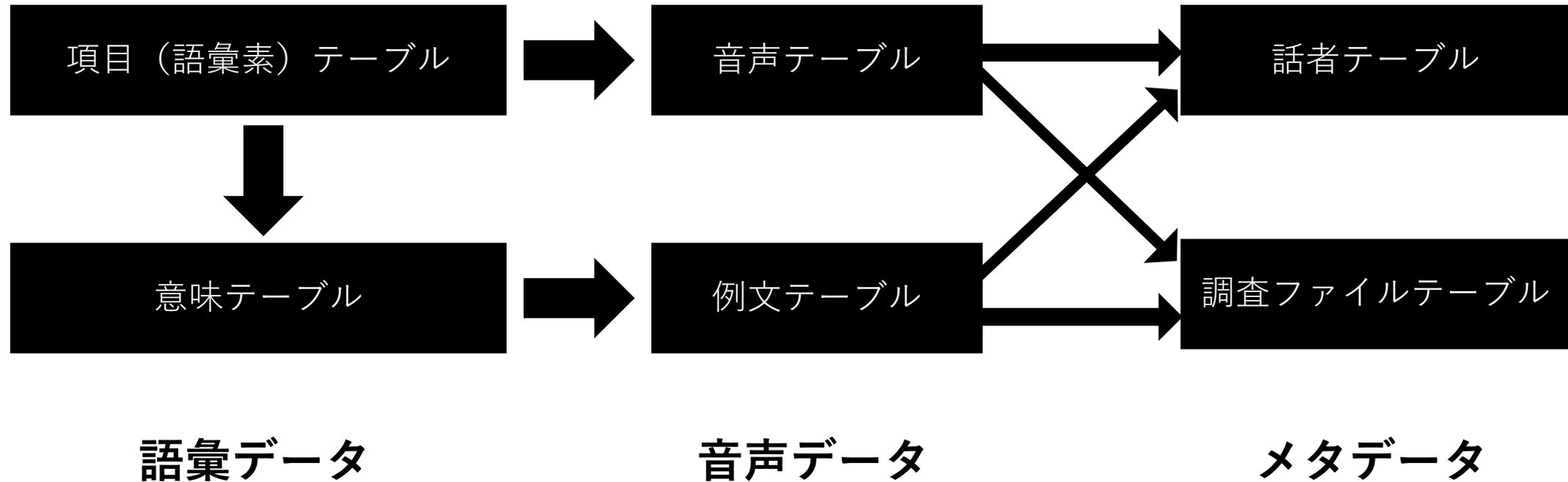
3. 日琉諸方言の語彙データを構築する

- 「語彙データフォーマット（仮）」
 - 最上位の記述単位：（凡そ）一地点の語彙素
 - 属性：一般的な琉和辞典＋音声（単独、ア資料、例文）

テーブル	単位	分類
項目	語彙素	語彙データ
意味	個別意味	語彙データ
音声	単独・ア資料ファイル	音声データ
例文	個別例文	音声データ
話者	話者	メタデータ
調査ファイル	調査ファイル	メタデータ

3. 日琉諸方言の語彙データを構築する

- 「語彙データフォーマット（仮）」



3. 日琉諸方言の語彙データを構築する

- 「語彙データフォーマット（仮）」
 - 編集作業において
 - A. 「項目テーブル」と「意味テーブル」を分けていない
 - B. 「単独音声テーブル」と「アクセント」を分ていけない
 - A. は編集便宜のため
 - B. は単独発音はア資料の一種とみなせるから（＝性質が近い）

項目兼意味テーブル

欄名	タイプ	記述	重複
項目Id	数字	項目ID	✓
順	数字	意味番号	
意味id	数字	意味ID	
表記1	文字	アクセント記号付き仮名表記	✓
表記2	文字	アクセント記号無し仮名表記	✓
IPA	文字	ローマ字表記	✓
拍数	数字	項目の拍数	✓
型	文字	項目のアクセント情報	✓
品ID	数字	項目の品詞ID	✓
品詞	文字	項目の品詞（略号）	✓
分類	文字	意味の分類範疇（「魚」「地名」）	
活用	文字	項目の活用クラス	✓
パラ	文字	動詞などの代表活用形	✓
意味	文字	意味記述の相当語	
ルビ	文字	意味記述の相当語のルビ	
読み	文字	意味記述の相当語の読み	
解説	文字	意味記述の解説	
備考	文字	項目に関する備考	✓
同	数字	異形態番号	✓
類	数字	類義語番号	✓

項目兼意味テーブル

- 項目の属性は**重複**
- 意味番号は「順」で代表できない
- 品詞IDはエクセルのソート機能の欠点の解消（つまり、編集用）

欄名	タイプ	記述	重複
項目Id	数字	項目ID	✓
順	数字	意味番号	
意味id	数字	意味ID	
表記1	文字	アクセント記号付き仮名表記	✓
表記2	文字	アクセント記号無し仮名表記	✓
IPA	文字	ローマ字表記	✓
拍数	数字	項目の拍数	✓
型	文字	項目のアクセント情報	✓
品ID	数字	項目の品詞ID	✓
品詞	文字	項目の品詞（略号）	✓
分類	文字	意味の分類範疇（「魚」「地名」）	
活用	文字	項目の活用クラス	✓
パラ	文字	動詞などの代表活用形	✓
意味	文字	意味記述の相当語	
ルビ	文字	意味記述の相当語のルビ	
読み	文字	意味記述の相当語の読み	
解説	文字	意味記述の解説	
備考	文字	項目に関する備考	✓
同	数字	異形態番号	✓
類	数字	類義語番号	✓

項目兼意味テーブル

- データのほかに可視化用の欄も
- 「読み」は検索・逆引き作成用
(「読み」は「相当語」の曖昧性
解消のためでもある)

「読み」があれば、逆引きは自動的に作成可能

欄名	タイプ	記述	重複
項目Id	数字	項目ID	✓
順	数字	意味番号	
意味id	数字	意味ID	
表記1	文字	アクセント記号付き仮名表記	✓
表記2	文字	アクセント記号無し仮名表記	✓
IPA	文字	ローマ字表記	✓
拍数	数字	項目の拍数	✓
型	文字	項目のアクセント情報	✓
品ID	数字	項目の品詞ID	✓
品詞	文字	項目の品詞 (略号)	✓
分類	文字	意味の分類範疇 (「魚」 「地名」)	
活用	文字	項目の活用クラス	✓
パラ	文字	動詞などの代表活用形	✓
意味	文字	意味記述の相当語	
ルビ	文字	意味記述の相当語のルビ	
読み	文字	意味記述の相当語の読み	
解説	文字	意味記述の解説	
備考	文字	項目に関する備考	✓
同	数字	異形態番号	✓
類	数字	類義語番号	✓

項目兼意味テーブル

- 「相当語」と「解説」と「備考」
- 「相当語」は対象言語の訳語
- 「解説」は意味を解説する文章
- 「備考」は項目（語彙素）に関する情報

欄名	タイプ	記述	重複
項目Id	数字	項目ID	✓
順	数字	意味番号	
意味id	数字	意味ID	
表記1	文字	アクセント記号付き仮名表記	✓
表記2	文字	アクセント記号無し仮名表記	✓
IPA	文字	ローマ字表記	✓
拍数	数字	項目の拍数	✓
型	文字	項目のアクセント情報	✓
品ID	数字	項目の品詞ID	✓
品詞	文字	項目の品詞（略号）	✓
分類	文字	意味の分類範疇（「魚」「地名」）	
活用	文字	項目の活用クラス	✓
パラ	文字	動詞などの代表活用形	✓
意味	文字	意味記述の相当語	
ルビ	文字	意味記述の相当語のルビ	
読み	文字	意味記述の相当語の読み	
解説	文字	意味記述の解説	
備考	文字	項目に関する備考	✓
同	数字	異形態番号	✓
類	数字	類義語番号	✓

「相当語」と「解説」を分けるのは極めて重要！

項目兼意味テーブル

- 「相当語」は構造を持たせている
「(ヘッド)相当語1。(ヘッド)相当語2」
- ヘッドは曖昧性解消のため

欄名	タイプ	記述	重複
項目Id	数字	項目ID	✓
順	数字	意味番号	
意味id	数字	意味ID	
表記1	文字	アクセント記号付き仮名表記	✓
表記2	文字	アクセント記号無し仮名表記	✓
IPA	文字	ローマ字表記	✓
拍数	数字	項目の拍数	✓
型	文字	項目のアクセント情報	✓
品ID	数字	項目の品詞ID	✓
品詞	文字	項目の品詞(略号)	✓
分類	文字	意味の分類範疇(「魚」「地名」)	
活用	文字	項目の活用クラス	✓
パラ	文字	動詞などの代表活用形	✓
意味	文字	意味記述の相当語	
ルビ	文字	意味記述の相当語のルビ	
読み	文字	意味記述の相当語の読み	
解説	文字	意味記述の解説	
備考	文字	項目に関する備考	✓
同	数字	異形態番号	✓
類	数字	類義語番号	✓

音声テーブル

- 単位：単独とア資料の音声ファイル
- 「*」の属性は自動的に派生（実習参照）

欄名	タイプ	記述
項目ID*	数字	項目ID
話者*	文字	話者略号（またはID）
タイプ*	文字	音声のタイプ：「X」は単独発音、「X+」はアクセント資料
粹文*	文字	アクセント資料の粹文
書起し	文字	該当音声の書き起こし
（訳）	（文字）	（書き起こしの訳※）
型	文字	該当音声のピッチ実現に基づいたアクセント型
備考	文字	該当音声に関する備考
調査ファイル*	文字	該当音声抽出された元の調査ファイル（名称かID）
 ファイル名*	文字	音声のファイル名

※語彙DBには必要だが、現在は「粹分」で代用

音声テーブル

- **項目ID**で紐づけられ、項目にぶら下がっている
- **メタ情報** (2種類)

欄名	タイプ	記述
項目ID	数字	項目ID
話者	文字	話者略号 (またはID)
タイプ	文字	音声のタイプ: 「X」は単独発音、「X+」はアクセント資料
粹文	文字	アクセント資料の粹文
書起し	文字	該当音声の書き起こし
(訳)	(文字)	(書き起こしの訳)
型	文字	該当音声のピッチ実現に基づいたアクセント型
備考	文字	該当音声に関する備考
調査ファイル	文字	該当音声抽出された元の調査ファイル (名称かID)
ファイル名	文字	音声のファイル名

音声テーブル

- タイプ記述
- 言語学的処理

欄名	タイプ	記述
項目ID	数字	項目ID
話者	文字	話者略号（またはID）
タイプ	文字	音声のタイプ：「X」は単独発音、「X+」はアクセント資料
枠文	文字	アクセント資料の枠文
書起し	文字	該当音声の書き起こし
（訳）	（文字）	（書き起こしの訳）
型	文字	該当音声のピッチ実現に基づいたアクセント型
備考	文字	該当音声に関する備考
調査ファイル	文字	該当音声抽出された元の調査ファイル（名称かID）
ファイル名	文字	音声のファイル名

4. 「語彙 DB」の概要

- RDB

- 基本的に「語彙データフォーマット（仮）」の構造を踏襲
 - => 「フォーマット」が満たしている要件をすべて継承している
 - => 互換性（別のフォーマットでもエクスポート可）

- <https://kikigengo.ninjal.ac.jp/data/tango/search> で2022年9月より公開

4. 「語彙 DB」の概要

- RDB

- 基本的に「語彙データフォーマット（仮）」を踏襲
- 何点かの違い
 - 「項目テーブル」と「意味テーブル」を分けている
 - 「音声テーブル」と「アクセント資料テーブル」を分けている
 - 英語版用の属性が加えられている※
 - メタデータは未対応だが、3月ごろ対応する予定

(※英語対応は必須でない)

4. 「語彙 DB」の概要

- 語彙DBでできること
 - 複雑な語彙データの登録
 - 上限なしの音声（単独、ア資料、例文）の登録
 - アクセント資料が登録可能（=>ア資料のDBとしても利用可能）
 - 英語版掲載可能

4. 「語彙 DB」の概要

- 内容（掲載済みと掲載予定の一部）

データ	項目数	音声数	掲載
前期DBの内容	10,000	10,000	済
『みんなふつ語彙集』	6,000	19,000	済（音声は一部）
『多良間方言辞典』	16,000	50,000	予定（3月まで）
『どうなんむぬい辞典』	2,000	nr	予定（3月まで）
『及位方言辞典』	8,000	Nr	予定（3月まで）
『池間方言辞典』（ア資料）	6,000	20,000?	予定（3月～5月）
『波照間方言小辞典』	4,000	10,000	予定（4月）
前期PJ共同研究員提出分	数千	数千	予定（4月～）
南琉球セリック分	数万	数万	予定（4月～）

5. 「語彙 DB」の今後の予定

- これよりシステムの改善・改変を重ねていく。
 - 改善・改変1が発注済みで、本月着工
 - 検索機能の強化
 - メタデータの導入（データ出典・話者・調査ファイル）
 - 同言語IDの導入

5. 「語彙 DB」の今後の予定

- これよりシステムの改善・改変を重ねていく。
 - 改善・改変1が発注済みで、本月着工、3月納品
 - 検索機能の強化
 - メタデータの導入 (データ出典・話者・調査ファイル)
 - 同言語IDの導入



語彙データ整備の「業績化」を重視する

5. 「語彙 DB」の今後の予定

- 同源語IDに伴う資源化（通時的観点の導入）
 - Celik et al (submitted)で試案され、「琉球祖語DB」に使用している同言語IDを導入
 - 同言語IDを導入することにより、他の語彙資源との「**接続**」が可能になる
 - 「琉球祖語DB」
 - Martin (1987)
 - アクセント辞典類
 - Unidic
 - 『日国』
- 等々

6. 語彙 DB への登録

- 構想

- 公表されているデータを可視化する
- (例外を除き) 基本的に公表済みのデータをDB登録

(共同研究員分含め) 語彙データを語彙資料として投稿し、作成者の1業績として公表し、学術レポジトリ登録が完了してから、DBへ登録する

=> 五十嵐・山田PJの共同で『日本列島の諸言語の研究資料シリーズ』電子雑誌を創刊

6. 語彙 DB への登録

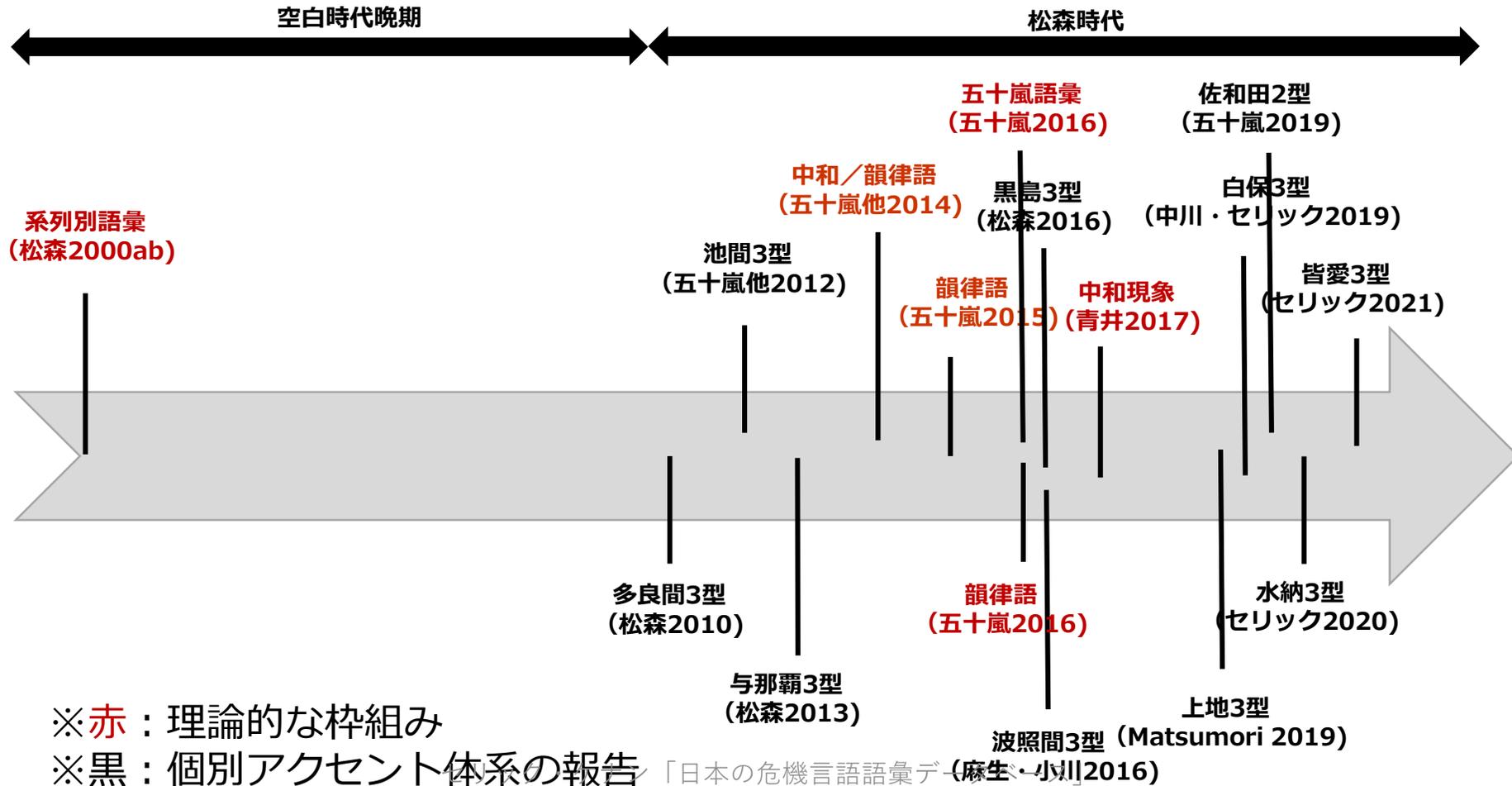
- 登録についてはセリック（国立国語研究所）
kcelik@ninjal.ac.jp まで連絡

7. アクセント資料の活用

- 南琉球のアクセント研究の動向と問題点

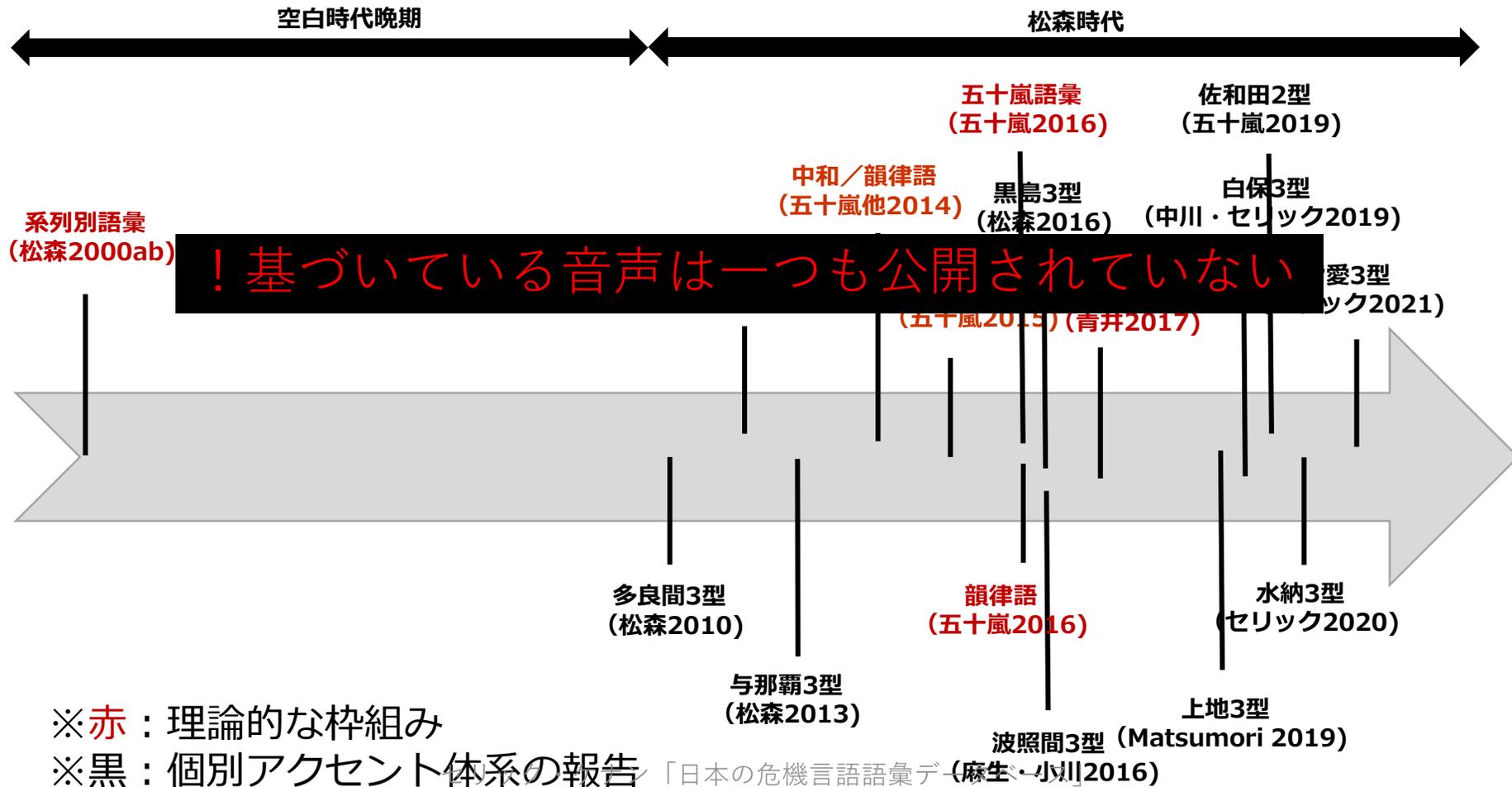
7. アクセント資料の活用

□ 近年、猛烈な発展（理論、報告）



7. アクセント資料の活用

□ 近年、猛烈な発展（理論、報告）



7. アクセント資料の活用

□ 近年、猛烈な発展（所属）

	年	方言	数（凡そ）
『石垣方言辞典』	2003	石垣	15,000
「与那国方言のアクセント資料」	2011~2016	与那国	数千
『竹富方言』	2013	竹富	<800
『多良間方言辞典』	2020	多良間仲筋	15,000
『鳩間方言辞典』	2020	鳩間	20,000
『みんなふつ語彙集』	2022	水納	5,000
『池間方言辞典』	2022	池間西原	>3,000

7. アクセント資料の活用

□ 近年、猛烈な発展（所属）

	年	方言	数（凡そ）
『石垣方言辞典』	2003	石垣	15,000
「与那国方言のアクセント資料」	2011~2016	与那国	数千
『竹富方言』	2013	竹富	<800
『多良間方言辞典』	2020	多良間仲筋	15,000
『鳩間方言辞典』	2020	鳩間	20,000
『みんなふつ語彙集』	2022	水納	5,000
『池間方言辞典』	2022	池間西原	>3,000

！現時点では2件のみ

7. アクセント資料の活用

- 南琉球のアクセント研究の動向と問題点
- 単独の発音だけではアクセント型が分からない
 - => アクセント資料（対象語を枠文に入れた資料）が必須
- 認定がかなり困難な方言
 - => 認定の精度に影響を与え得る
- 音声の公開が望ましいが、そうされていない
 - => 構造的な要因

7. アクセント資料の活用

- 南琉球のアクセント研究の動向と問題点
- ローレンス (2013)
 - 「『竹富方言辞典』には 0 (平板型音調) とあるが、起伏型である。」
 - 「『竹富方言辞典』には 2 (起伏型音調) とあるが、平板型である」
- 何に起因するずれが不明
 - 発音間違い、認定間違い、記入間違い、個人差など、知りようがない

参考文献（1節-6節）

- 上野善道(2016)「琉球与那国方言体言のアクセント資料(5)」『琉球の方言』40: 71-105.
- かりまたしげひさ(2005)「「琉球語音声データベース」一消滅に瀕する琉球語の記録・保存一」『音声研究』9-3: 23-28.
- カルリノ・サルバトーレ(2022)「「日琉諸語オンライン辞書」の紹介」『日本語の研究』18-3: 52-59.
- Celik, K., Nakazawa, K., Aso, R. (submitted) A Database of Proto-Ryukyuan.

参考文献（7節）

- 青井隼人 (2012) 「宮古多良間方言における「中舌母音」の音声的解釈」『言語研究』142: 77-94.
- 青井隼人 (2017) 「南琉球宮古多良間方言における2種類のアクセント型の中和」『国立国語研究所論集』13: 1-23.
- 麻生玲子・小川晋史 (2016) 「南琉球八重山語波照間方言の三型アクセント」『言語研究』150: 87-115.
- セリック・ケナン (2020) 『南琉球宮古語史』博士論文京都大学.
- セリック・ケナン (2020) 「南琉球宮古語水納島方言のアクセント体系と基礎語彙」『琉球の方言』44, 243-281.
- セリック・ケナン (2022) 「下地皆愛方言のアクセント体系に関する予備的報告」『言語記述論集』13, 215-290.
- Davis, Christopher (2018) 「沖縄県竹富町小浜島・八重山語小浜言葉」琉球大学国際沖縄研究所(編)『シマジマのしまくとぅば: 平成29年度危機的な状況にある言語・方言のアーカイブ化を想定した実地調査研究: 文化庁委託事業報告書』181-199. 沖縄: 琉球大学国際沖縄研究所.
- 五十嵐陽介 (2015) 「南琉球宮古語多良間方言のアクセント型の記述」『比較日本文学研究』8: 1-42.
- 五十嵐陽介 (2016a) 「南琉球宮古語池間方言・多良間方言の韻律構造」『言語研究』150: 33-57.
- 五十嵐陽介 (2016b) 「アクセント型に対応に基づいて日琉祖語を再建するための語彙リスト」『日琉語類別語彙』』『日本語学会2016年度春季大会予稿集』233-238.
- 五十嵐陽介 (2019) 「琉球宮古語伊良部佐和田方言のアクセント体系は三型あるいは二型であって一型ではない」(一橋大学五十嵐ゼミ令和元年夏期「終日ゼミ」発表原稿) 1-14.
- 五十嵐陽介・田窪行剛・林由華・ペラルル・トマ・久保智之 (2012) 「琉球宮古語の池間方言のアクセント体系は三型であって二型ではない」『音声研究』16(1): 134-148.
- Igarashi, Yosuke, Takubo, Yukinori, Hayashi, Yuka and Kubo, Tomoyuki (2018) "Tonal Neutralization in the Ikema Dialect of Miyako Ryukyuan". Tonal Change and Neutralization, edited by Haruo Kubozono and Mikio Giriko, Berlin, Boston: De Gruyter Mouton, 83-128. <https://doi.org/10.1515/9783110567502-005>
- 金田一春彦 (1974) 『国語アクセントの史的研究-原理と方法』東京: 塙書房.
- 平山輝男・大島一郎・中本正智 (1967) 『琉球先島方言の総合的研究』東京: 明治書院.
- ローレンス・ウエイン (2000) 「八重山方言の区画について」石垣繁 (編)『宮良當社記念論集』547-559. 沖縄: 宮良當社生誕百年記念事業期成会.
- ローレンス・ウエイン (2013) 「竹富島方言アクセントと『系列別語彙』: 附 竹富島方言版「北風と太陽」」『琉球の方言』37:1-24
- 松森晶子 (2000a) 「琉球の多型アクセント体系についての一考察—琉球祖語における類別語彙3拍語の合流の仕方—」『国語学』51(1): 93-108.
- 松森晶子 (2000b) 「琉球アクセント調査のための類別語彙の開発—沖永良部島の調査から」『音声研究』4(1): 61-71.
- 松森晶子 (2010) 「多良間島の3型アクセントと『系列別語彙』」上野晋道 (監)『日本語研究の12章』490-503. 東京: 明治書院.
- 松森晶子 (2013) 「宮古島における3型アクセント体系の発見: 与那覇方言の場合」『国立国語研究所論集』6: 67-92.
- 松森晶子 (2015) 「南琉球の三型アクセント体系: その韻律単位に関する考察」『日本女子大学紀要文学部』64: 55-92.
- 松森晶子 (2016) 「八重山語黒島方言アクセントの仕組み: その韻律範疇PWd と下がり目の出現条件」『言語研究』150: 59-85.
- Matsumori, Akiko (2019) "Prosodic Unit, Recursive Structure, and Nature of Accent in Miyako Ryukyuan". The Linguistic Review, 36(1), 51-83. <https://doi.org/10.1515/1tr-2018-2007>
- 中川奈津子・セリックケナン (2019) 「琉球八重山白保方言のアクセント体系は三型であって、二型ではない」『日本語学会2019 年度春季大会予稿集』89-96.