e-Learning教育学会 外国語教育のための I C T 無料ワークショップ 2019年3月31日 OIT 梅田 タワー





hinoyoshihiro@gmail.com http://www.hinox.org

CTAの設置アドレス Chinese Text Analyzer





CTAの設計意図

中国語で書かれたテキストを分割したい。 我家离大学不太远。

→ 我/家/离/大学/不/太/远/。/

分割した中国語をピンインに変換したい。

→ Wǒ jiā lí dàxué bú tài yuǎn.

- 分割した結果を利用したい。
 - → 単語リストの作成、単語の集計etc…
- 教科書コーパスを作りたい。

→ 品詞情報を付与した教材コーパスの構築

授業準備を、少しでも手抜きしたい。





入門学習者向けの教材の作成補助

・単語の抜き出し・ピンイン(発音符号)の付与

テキストデータの解析

• 含まれる単語の量や難易度の解析

教科書コーパス構築の補助

例文の検索



CTA開発の背景

- 中国語のテキストは区切りがない。
- 既存のピンイン入力は面倒。
- ピンインで入力するが、ピンインを入力しない。
- •既存サイトやプログラムの限界。



システム開発の背景

- 天气好的时候,我常常步行 来大学。今天我就是走着来的。今天早上我是8点出门的。
 时间还早,我顺便走进了路边的一家书店。书店的书架上摆着很多书和杂志,我买了一本新出的漫画。
- 中国語は単語の切れ目が分からない。

Lesson 7 電話で友達と約束 (~がしたい	·)
(Stoy) 張君は来週の土曜日に鈴木さんを映画に誘います。	126
张明:你下个星期六有空儿吗? Nī xià ge xīngqīliù yǒu kòngr ma?	新出単語 「下个 xià ge 次の
徐木:下个星期六是几号? Xià ge xingqiliù shì ji hào?	□ 星期六 xīngqīliù 上曜日 □ 空儿 kòngr ひま □ 号 hào ~日
张明:十八 号。 Shíbā hào.	□ 怎么∫zēnmele どうしたの □ 想 xiāng ~したい □ 看 kān 見る
^{鈴木:} 有空儿。怎么了? Yǒu kòngr. Zénme le?	□ Ц影 diànyīng 映画 □ 点 diàn ~啊 □ 下午 yiànyī 午後
张明:我想看电影,一起去吗? Wó xiáng kàn diànyǐng, yìqí qù ma?	 好啊 hāo a いいですよ 那 nà それじゃあ
鈴木: 几 点 的 电影? Jī diān de diànyíng?	□中 ban 30万、~中 □见 jiàn 会う
张明:下午两点。 Xiàwù liàng diàn.	
於木:好啊,那我们一点半见吧。 Hão a, nà wǒmen yì diǎn bàn jiàn ba.	

システム開発の背景	
•我家离大学不太远,可以坐 7 電話で友達と	と約束 (~がしたい)
电午、 行。天 ^{张明} : 你下个星期六有空儿吗? 步行录 Ni xià ge xīngqīliù yǒu kòngr ma?	
着来的 Art 下个 星期六 是 几 号? 门的。 Xià ge xīngqīliù shì jǐ hào?	
进了路 新明:十八 号。 的书杂 Shíbā hào.	空儿 吗? kôngr ma? 号? hào? か 新出単語 ♪ 127 ♪ 127 ♪ 127 ↓ 12
→ 単語の 本: 有空儿。 怎么 了? Yǒu kòngr. Zěnme le?	□ 想 xiáng ~したい □ 看 kàn 見る □ 电影 dianying 映画 □ 広 dian ~時 □ 下午 xiawū 午後 起 去 吗? □ が啊 háo a いいですよ □ 」 gù ma? □ 郡 nà それにゃあ
张明:我想看电影,一起去吗? Wǒ xiǎng kàn diànyǐng, yìqǐ qù ma?	□毕bàn 30分、~# □见 jiàn ☆う 一点 半 见 吧。 yì diǎn bàn jiàn ba.



Winのピンイン入力

ChineseWriter



全パターンを身近に置いてコピペ Āáǎà / ēéěè / ōóǎò / īíǐì / ūúǎù / ūúǚüüü

(画像引用) https://www.kodensha.jp/index/wp-content/uploads/2018/03/chinesewriter11.pdf



Macのピンイン入力

「声調記号」を入力し、「アルファベット」する。

設定
 「環境設定」から「言語とテキスト」を選択
 「入力ソース」から「U.S. Extended」を探して有効にする。

• 入力方法

ピンインを入力する際にはIMEを「U.S. Extended」に切り替える。

1声:Option + 「a」 十母音 2声:Option + 「e」 十母音 3声:Option + 「v」 十母音 4声:Option + 「_」 十母音



変換サービス

参考)どんとこい中国語



wõ	jiā	lí	dà	xué	bú,	/bù	tà	i yu	ıăn	1	ł	kě y	уĭ	zu	iò d	liàr	n c	hē		qí	zì
我	家	离	大	学	7	不	太		远	,	1	可	以	샾	Ł	电	3	车		骑	自
háns	s/xír	ng/ł	nàna	s/hé	ng (chē		yě	kě	уÌ	ík	bù l	hán	g/x	ína	:/hà	ing	Xé	ng		
		行				车	,	也	可	以		步			ł	行					
tiār	n qì	hă	io/ł	nào	de/	dí/d	dì s	shí	hòu	1	V	۷Ŏ	ch	iáne	; ch	iána	s b	ù			
天	气		好			的		时	候	,		我	1	常	1	常	킛	F			
háns	s/×ír	ng/ł	nàna	s/hé	ng	lái	dà	xué	ś	j	īn	tia	ān	wŏ	ji	ù s	shì	zċ	ðu		
		行	-			来	大	学	•	~	今	天	5	我	勍	Ì.	是	쿳	ŧ		
zhe/	'zhuć	5/zł	náo	lái	de,	/dí/	/dì		jīn	n t	iār	n z	ã o	sh	àne	w	õ	shì		di	ăn
	着			来		的		0	今		天		早		F	ŧ	戈	是	8	ķ	ξ.
chū	mén	de,	/dí/	/dì		shí	jia	in/j	iàn	h h	ái/	/hua	án	ză	0	Y	۷Õ	sh	iùn	bià	in
出	IJ		的		•	时		间			÷	还		早	. ,		我	ļ	页	便	Ī
zŏι	i jìr	n le	e/1	ião	lù	biā	ān d	de/d	dí/c	lì :	уī	jia	ās	hū	dià	n		shū	ı d	iàn	
走	进		Ţ	•	路	达	2	É	的		-	家	2 -	书	店		,	书	,	店	
de/c	lí/dì	sł	hū .	iià	shài	ng b	ă	zł	ne/z	hu	6/2	zhá	o h	iěn	duā	sh	hū				
É	的	1	붠	架	F		摆			着	F		;	很	多	+	Ĵ				
hé/ł	ne/hu	lo/ł	hè/ł	nú/h	uó/ł	huò	zá	zhì		W	õ	mã	i i	le/	114	ăо	УĪ	bě	in b	ĸīn	chi
			和				杂	志	,	TH S	戝	굣	Ę		了			本	2	新	出
de/c	lí/dì	mà	àn ł	nuà																	
É	肑	ł		画	•																

変調

漢字?単語?



解決しないといけない問題

• 分割方法の構築

分割単位はー形態素?単語?フレーズ? 桌子/上/有/一/本/书/。/ 桌子上/有/一/本/书/。/ 桌子上/有/一本书/。/ ピンイン(多音字・変調・大文字)の処理 银行 (yínháng) 行为 (xíngwéi) 不去(bú qù) 不来(bù lái) 中国(Zhōngguó) 我去吃饭(Wǒ chī fàn.)



中国語テキストのサンプル

我家离大学不太远,可以坐电车、骑自行车,也可以步行。天气好的时候,我常常步行来大学。今天我就是走着来的。今天早上我是8点出门的。时间还早,我顺便走进了路边的一家书店。书店的书架上摆着很多书和杂志,我买了一本新出的漫画。

『スリム版中文在線基礎編』第9課



自分で分割

 我/家/离/大学/不/太/远,可以/坐/电车、骑/自行车,也/可以/步行。天气/好/的/时候,我常常/步行/来/大学。今天/我/就/是/ 走着/来/的。今天/早上/我/是/8/点/出门/的。时间还/早,我/顺便/走进了/路边/的/一/家/书店。书店/的/书架上/摆着/很/多/ 书/和/杂志,我买/了/一/本/新出/的/漫画。





ブラウザでさくっとするためのCTAの設計図

- 単語の分割をスムーズに行うために形態素解析システム

MeCabと独自の解析用辞書を利用。

- ブラウザ(WEB)で処理するためにJavaで実装された形態素解析 器lgo-phpを利用。
- ブラウザで利用でき、修正も簡単な操作でできる。スマホで空き 時間にさくっと素材を加工できる→タップで操作できるようにデ ザイン。
- 分割したデータを二次利用しやすいように様々な形で提供。



CTA

中国語の形態素解析 行い単語に分割

■分割した単位でピン・ ンに変換

分割修正が可能

■変調も解析辞書に基* き処理



このサイトは現在構築途中のものをβ版として公開しています。そのためバグや気になる点がありましたらこ ちらから投稿お願いします。また使用方法やこのサイトでできることはこちらをご確認ください。

我家离大学不太远,可以坐电车、骑自行车,也可以步行。天气好的时候, 我常常步行来大学。今天我就是走着来的。今天早上我是8点出门的。时间还 早,我顺便走进了路边的一家书店。书店的书架上摆着很多书和杂志,我买 了一本新出的漫画。

109文字

分割辞書の選択

標準辞書

漢字+ピンイン 漢字 ビンイン

エクスポート

「漢字+ピンインモード」では複数の読み方がある漢字は青色で表示されます。クリックして正しいピン インを選択してください。分割が間違っている場合は「分割修正」をクリックして修正してください。

我	家	函	大学	不	太	远	,	可以
Wŏ	jiā	li	dàxué	bù	tài	yuăn	,	kěyĭ
电车	2	骑	自行车	,	也	可以	步行	۰
diànchē	,	qí	zixingchē	,	yĕ	kěyĭ	bùxíng	2
好	的	时候	,	我	常常	步行	来	大学
hăo	de	shíhòu	,	wŏ	chángcháng	bùxing	lái	dàxué
今天	我	就是	走	着	来	的		今天
Jīntiān	wŏ	jiùshì	zŏu	zhe	lái	de	•	Jīntiān
我	是	8	占	L1 出	的	-	时间	还



①分析したいテキストを入力する

Chinese Text Analyser β版 使い方 このサイトについて LINK-

このサイトは現在構築途中のものをβ版として公開しています。そのためバグや気になる点があり ましたらこちらから投稿お願いします。また使用方法やこのサイトでできることはこちらをご確認 ください。

我家离大学不太远,可以坐电车、骑自行车,也可以步行。天气好的时候, 我常常步行来大学。今天我就是走着来的。今天早上我是8点出门的。时间还 早,我顺便走进了路边的一家书店。书店的书架上摆着很多书和杂志,我买 了一本新出的漫画。





②分割結果の表示





デフォルトでは 上下二段に分割した単語 とピンインが表示

この段階では 形態素解析に基づき 「単語」に分割

のべ語数: 69 異なり語数: 51

語数のカウント



漢字+ビ	ンイン	漢字 ピン	ノイン 分割修	E	ム案処理 🛛 エ	クスポート	印刷	クリア
「漢字+ インを選	ピンインモ 択してくた	ヨー ド」 マロ Eさい。 5	「分割処理	重」と 処理	と「ルー」 里	たわまま クリ レ」 ^{て修正}	ックして正 Eしてくだ	Eしいピン さい。
我	家	菡	大学	不	太	远	,	可以
Wŏ	jiā	lí	dàxué	bù	tài	yuǎn	,	kĕyĭ
电 <mark>车</mark>	•	骑	自行车	,	也	可以	步行	•
diànchē		qí	zixingchē		yĕ	kĕyĭ	bùxíng	
和	杂志	,			了	一本	新	出
hé	zázhi	,	多音字	ii	le	yīběn	xīn	chū
漫画	•							
1100 100 1000								

mànhuà .

*

のべ語数: 69 異なり語数: 51



*

CTAの修正機能の考え方





修正機能:ルールに基づく自動処理

単語分割≠ピンインルールでは必ずしもないため手動修正機能とルールに基づく自動変換機能を実装、必要な処理を「選んで」結果に反映させる

例)

(1)1文字の動詞で同じ動詞が続く場合は「くっつけて」「後ろの声調を消す」

 看 看 A 看看 kànkan

(2) 1文字の動詞+「一」+同じ動詞となる場合「全部くっつけて」
 「一の声調を消す」
 看 一 看 → 看一看 kànyikàn

修正機能②:単語の再分割

「分割修正」モードでは文字をクリックするとその文字の後ろが分割され,"/"をクリックすると前後の文字が結合される。たとえば下の文字列の中で赤で示した「/」をクリックすると前後の単語を結合、赤(下線)で示した漢字をクリックすると直後の漢字と分割される。



我/是/中国人。我/每天/<mark>来</mark>/大学。

修正機能③:ピンインの修正

「漢字+ピンイン」モードで修正できる

- 多音字の場合文字をクリックで変更できる

<u>zhòng</u> ⇔ <u>chóng</u>

多音字以外は声調符号をクリックすると、声調符号なしに変更できる

sh<u>à</u>ng ⇔ sh<u>a</u>ng

- 変調する"ー"や"不"もクリックで変更できる
- •r化は以下のようにクリックすると変更できる

ér⇔ r







漢字+ピンイン
 漢字
 ピンイン
 分割修正
 私案処理
 エクスポート
 印刷
 クリア
 「漢字+ピンインモード」
 では複数の読み方がある漢字は青色で表示されます。クリックして正しいピン
 インを選択してください。分割が間違っている場合は「分割修正」をクリックして修正してください。

我家离大学不太远,可以坐电车、骑自行车,也可以步行。天气好的时候,我常常步行来大学。今天我就是走着来的。今天早上我是8点出门的。时间还早,我顺便走进了路边的一家书店。书店的书架上摆着很多书和杂志,我买了一本新出的漫画。

のべ語数:67 異なり語数:49

完成したデータは、「漢字」モード 「ピンイン」モードに切り替えてコ ピーすることや、エキスポート機能 でいくつかの種類に加工することが できます。

「教材を作る」ための機能

エクスポート機能

 PDFエクスポート (単語学習プリント)
 <u>分割結果の</u>エクス

• 0

• 刀割記未のエクス ポート

単語の集計

- のべ語数と異なり語 数集計
- ・異なり語数の一覧取

 得

単語の照合

- 新旧HSKリストと
 の一致照合
- 中国語教育学会『
 中国語初級段階学
 習指導ガイドライン』「学習語彙
 表」



エクスポート機能



 分割結果のエクス ポート

エクスポート
この機能は構築中です。エクスボートの形式を選択してください。
1.出力するテキストの種類を指定する
ビンイン×漢字
2.改行場所を指定する
単語で改行
3.出力形式を指定する
コビー可能なテキスト
Wŏ
我
jiā
家
lí
<u>철</u>

□ 1	我	Wŏ	
□ 2	家	jiā	
□ 3	离	lí	
□ 4	大学	dàxué	
□ 5	不	bù	
□ 6	太	tài	
07	远	yuǎn	
□ 8	可以	kěyĭ	
□ 9	坐	zuò	
□ 10	电车	diànchē	
□ <u>11</u>	骑	qí	
□ 12	自行车	zìxíngchē	
□ 13	也	yě	
□ 14	可以	kěyĭ	
□ 15	步行	bùxíng	
□ 16	天气	Tiānqì	
□ <u>17</u>	好	hǎo	
□ 18	的	de	
□ 19	时候	shíhòu	
□ 20	我	wŏ	

単語の集計

Chinese	Text Analyser β版	異なり語	
		単語	出現回数
4	ドロの未可	我	6
	語数集計	的	6
•	異なり 語数の一覧 取得 远	大学	2
	学	可以	2
		步行	2
	一夜回	来	2
		今天	2
	109文字	家	2
	分割辞書の選択		2

単語の照合



LVが低い単語が暖色系、難易度があがると寒色系の色 になるように指定している。 また各LVごとに単語の集計をする機能がついている。

おまけの機能

ります。



ご清聴ありがとうございました。





氷野 善寛(目白大学) hinoyoshihiro@gmail.com

本研究の一部はJSPS科研費17K02956と16K16839の助成を受けたものです。

