Modified Authenticity:

A Sentence Corpus and Grammar Search Tool for L2 Beginners

Kiyomi CHUJO

Nihon University





Kathryn OGHIGIAN

Waseda University

Situation & Problem



report_unlweb.pdf

TOEIC 2014

Table 1: Mean Performance by Native Country/Region

Comments	Listo	ening	Rea	ding	Total		
Country/Region	Mean	(SD)*	Mean	(SD)*	Mean	(SD)	
CANADA	432	(69)	393	(82)	825	(146)	
GERMANY	425	(74)	362	(97)	787	(165)	
INDIA	408	(83)	361	(104)	769	(182)	
PORTUGAL	404	(85)	362	(100)	766	(180)	
LEBANON	409	(80)	354	(97)	763	(170)	
CZECH REPUBLIC	398	(97)	362	(103)	760	(194)	
BELGIUM	393	(93)	362	(100)	756	(187)	
ITALY	371	(75)	355	(79)	725	(146)	
FRANCE	377	(89)	342	(94)	719	(177)	
PHILIPPINES	384	(78)	325	(95)	710	(166)	
ARGENTINA	368	(113)	340	(114)	708	(222)	
TUNISIA	374	(87)	334	(92)	708	(172)	
COSTA RICA	378	(93)	326	(98)	704	(183)	
MOROCCO	375	(86)	324	(94)	699	(174)	
SPAIN	345	(101)	335	(100)	680	(195)	
SLOVAKIA	358	(111)	317	(121)	675	(228)	
CHINA, PEOPLE'S REPUBLIC	350	(90)	322	(99)	671	(180)	
REUNION	352	(103)	318	(105)	670	(202)	
RUSSIA	359	(106)	308	(115)	667	(215)	
MALAYSIA	362	(91)	294	(106)	656	(190)	
TURKEY	355	(95)	299	(107)	654	(194)	
COLOMBIA	343	(114)	304	(111)	647	(219)	
KOREA (ROK)	351	(86)	295	(97)	646	(175)	
ALGERIA	344	(99)	299	(100)	643	(194)	
SENEGAL	344	(87)	298	(86)	641	(166)	
BRAZIL	334	(115)	303	(114)	636	(223)	
COTE D'IVOIRE (IVORY COAST)	324	(96)	297	(95)	621	(183)	
CAMEROON	322	(92)	282	(91)	604	(174)	
PERU	307	(119)	279	(114)	585	(226)	
GREECE	321	(95)	247	(97)	568	(184)	
POLAND	315	(111)	243	(117)	559	(221)	
TAIWAN	294	(97)	242	(101)	536	(190)	
HONG KONG	297	(101)	235	(109)	532	(202)	
PAKISTAN	299	(136)	227	(149)	527	(278)	
JAPAN	284	(91)	228	(98)	512	(181)	
UNITED ARAB EMIRATES	298	(109)	183	(106)	481	(206)	
THAILAND	274	(102)	207	(96)	481	(192)	
VIETNAM	249	(90)	229	(97)	478	(180)	
CHILE	257	(115)	218	(108)	475	(215)	
MACAO	258	(90)	191	(90)	449	(172)	
MONGOLIA	247	(111)	189	(106)	436	(210)	
INDONESIA	234	(104)	188	(97)	421	(194)	
EGYPT	218	(107)	177	(97)	395	(197)	
ALBANIA	208	(113)	168	(104)	376	(213)	

*SD=Standard Deviation

Japanese students are failing to master basic grammar in junior and senior high school.



Chujo, K., Yokota, K., Hasegawa, S. & Nishigaki, C. 2012. Identifying the General English Proficiency and Distinct Grammar Proficiency of Remedial Learners.

Journal of the College of Industrial Technology, Nihon University 45: 43-54.

Nihon Keizai Shimbun 2015年7月22日

びびの現場から

中学の復習 大学の 定番

第4部入試改革前夜

物園」の読み仮名 ゴマンセミナー」 際大(大阪府守口 プレットに次々と 年生向け授業プ 4564を10等 OWEL SH 。6月上旬、 お作法。 明教授は「アカデミズム に基礎からやるしかな 高校進学率が98%に達 の対応にもがいている。 しと明快だ。 多くの大学が新人生へ 大学も全人時代を迎 寺子屋のよう

就職できず進学

をくないが、指定 内容は小学校レベ 問題に学生たちが

っていた。

空入学した男子学

は崩れつつある。 就職」という旧来の構図 は大学、そうでなければ と明かす。「学力上位層 生徒は大学に押し込む」 は一高率で就職できない

見読しましょう」と、 えましょう」 教科 智古は「友達の名前

一括の心得を手取り

ら屈託ない。併用す

しも多いから助か

)は「忘れちゃっ

関西のある公立高教諭

容として)大学水率とは 「be動詞は(教育内 2月、

は最近の大学生の英語

だがこのも自動詞、

教科書では。定番が

語教科書を手掛ける出版

に営業回りに行

新宿)の担当者

老舗の商製費

摘在し話題になった。

科学省が、子葉科学大学

葉県銚子市)にこんな指

制作。大手私 復習を中心に

多くの学校で値

我風振にも、

大学とは思えな

校生が入ってくるため あらゆるレベルの高



Possible Solution

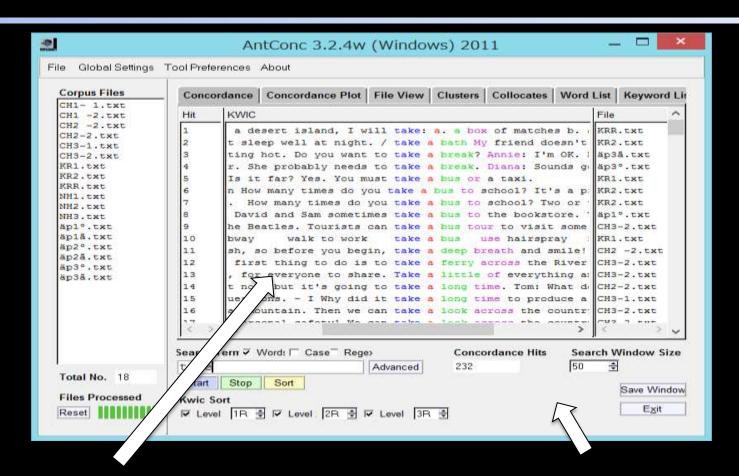
✓ L2

Grammar

✓ L2
vocabulary

nformation Technology, ABI, Computing Research Association, Microsoft, Google, and SAS. Th ship offers from IBM. The ACM Student Research Competition (SRC), sponsored by Microsoft etition (SRC), sponsored by Microsoft **Research**, continues to offer a unique forum for undergr aduate and graduate students to present their original research at well-known ACM-sponsore n contests held at 13 ACM conferences. Their **research** covered a range of computing innovation including computer science, IT and **research**, and engineering. ACM s Education Board continu designed to bring together the research and career interests of women in computing and hig nark) decision. The Internet **Research** Task Force (IRTF) urged the deployment of active queue ahtning strike. Since my job is **research** in immersive teleconferencing, I knew I had to dig int k (Internet network architect, currently senior research scientist at MIT) had warned about se king persistent queue while ignoring short-term bursts.7 Subsequent research tried to fix some widely exploited in both research papers and commercial virtualization systems. One useful ca a factor of two. Since then, further research, optimizations, and hardware acceleration have re ad would also be incurred for each sector. Much current research in I/O virtualization is focuse ing. I/O virtualization remains an active area of research and development in both academia a and electrical engineering departments at stanford university, his **research** interests include sy e relevant contextual information to trick specific victims. Academic research and commercial cs of computer security, Cormac Herley of Microsoft Research captured the problem succinctly ng email. A large body of research covers spam detection. However, research on detecting ph arch covers spam detection. However, research on detecting phishing email messages is span shing sites. Sheng et al. s research identified a gap between research and industry in terms of earch identified a gap between **research** and industry in terms of true positives. Academic res and industry in terms of true positives. Academic research has generally focused on heuristics w how to protect themselves. Two lines of research have sought to address these problems: 1

Data-Driven Learning (DDL)



Corpus (language data)



Corpus Tool (search software)

Reality of DDL

used mainly for advanced level learners

British N	National Corpus		Shogakukan Corpus Network
闭しる一使	家ワイントワヘ		ソート 語(句)集計表示変更 タワンロート へルブ
検索語(句	D): 表記形: There		
総件数:	317858 (2859 13 /	(1M語) 採用:3000 サブコーパス指定: 無	<< < 1 / 120 > >> 1 ページへ
3 A04	An art o	object is especially difficult for the critic , as there	are many ways in which its description can be approached.
4 A05		Then there	is the tale of a lying girl , as she may be , with whom he m
5 A06		lovers 'experience is drenched in imagery, there	is still a fine feeling of reality about the characters whic
6 A08	se always be cr	itics and there will always be oohs and ahs , there	will always be shit and there will always be vomit .
7 A08		Without images , he said , there	would not even be the wherewithal to talk about the death of
8 A08		Because in the imagination there	is no friction .
9 A08		What is there	is there .
10 A0D		There	was a grudging , muted agreement , Cissy and Freddy too preo
11 A0D		There	was nothing in his face to show how much he enjoyed teasing
12 A0F		When I arrived there	, there was a train for Harwich just about to leave .
13 A0F	n London so I :	got on a train and ended up here ; that 's all there	is to it , really . &equo
14 A0G		Finally , there	are large bush trees and trees of orchard size which are rea
15 A0H		This is because if there	is a person on both wing-tips , neither may have a good grip
16 A0H		Do n't try to go out backwards unless there	is ample space for the glider to go straight through .
17 A0H	round looping acc	idents and these must be analysed to see if there	are new lessons to be learned .
18 A0H		When there	is an opportunity during training for the pilot to decide wh
19 A0H	If the wi	ing-tip touches long grass or rough ground , there	will certainly be a bad swing and the launch will have to be
20 A0H	udder, it spee	eds up the outer wing-tip, creating more lift there	, and gives the inner wing &bquo sweep back &equo in_relat
21 A0H	not having tak	en that extra piece of lift some miles back , there	can be a tendency to become apathetic and not even try to pr
22 A0K	ophy and pull th	e marginal mover in from the periphery; for there	is a boundary beyond which the pilgrim can not be allowed to
23 A0L		Look , there	's lot of things I just ca n't handle . &e quo;
24 A0L		I lay as_if you were there	and loved the shape your back had made , had a sense of your
25 A0M		There	are some people who claim to be able to control an axe kick

Our Target Learners: Remedial Level

Who university students scoring TOEIC 300 or below, or CEFR A1, A2 level

What to improve understanding of basic grammar items taught in secondary school

How using DDL (Weber, 2001; Takanashi, 2009; Leńko-Szymańska & Boulton, 2015)

Using DDL for Remedial Learners

First Challenge lack of an appropriate level corpus

Forem insumbojor sit amet, consectetur abiniscing ejit, seddo eiusmod tempor incidibunt ut japore et bojore magna ajioua ut enim abinimim veniam, ouis mostrub exercitation ujijamco japorishisi ut ajioun ex ea commodio consecuat bius aute inurebojor in reprehenderit in vojuptate vejit esse ciliumbojore eu fugiatnujia pariatur excepteur sint occaecat cupidiatatnon projent, sunt in curpa oui officiadeserunt moditi anim id



Using DDL for Remedial Learners

Second Challenge
lack of an easy (and free) to use
search corpus tool



New Corpus + Tool

Sentence Corpus of Remedial English SCoRE

That was the man who stopped the train.

The man who just called was my brother.

The man who is playing the guitar is from Ghana.

Did you know that man who just left?

I got a letter from a man who lives in Palau.

The first man who walked on the moon was Neil Armstrong.

The last man who left was the fire chief.

Do you know that man who is waving at us?

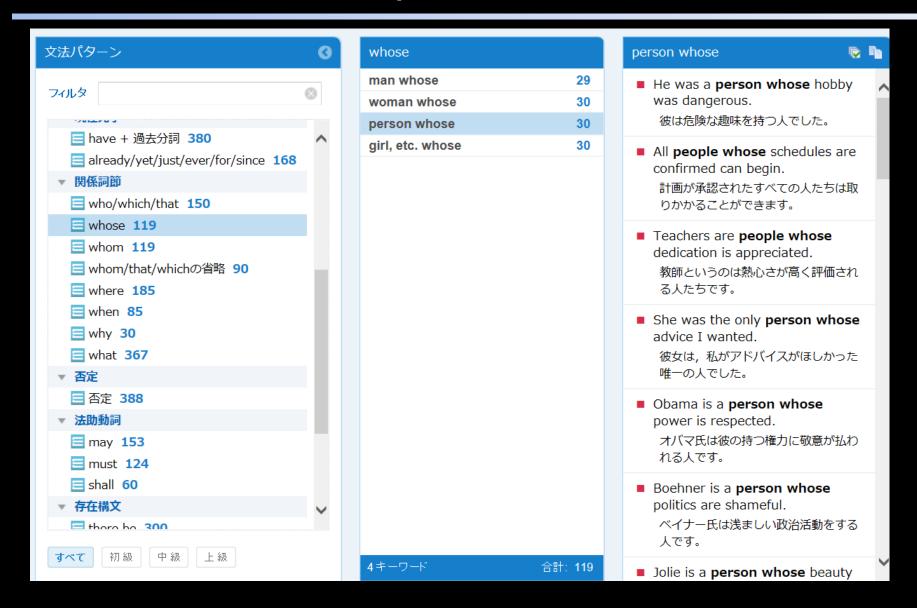
I want to introduce you to the man who wrote this book.

The old man who caused the accident died yesterday.

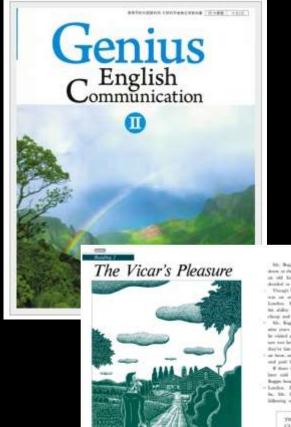
Steve Jobs was not the first man who created the personal computer, but he might be the most famous.

In the U.S., President Lincoln is best remembered as the man who ended slavery.

SCoRE (corpus) + GPPS (tool)



Step 1: Measuring Target Level



= up to U.S. school grades 4 and 5

At Buggs, diving his on year a lift, belief down at the remembels in four of him. He new on will furnishese that looked grienting, and depicts or may by for a visa.

Finish to see disposed on a star. We Register on an entry distinct the tensor day in London. We not seek to seek to the benefit of the seek of the see

Mis Baggio had a extent. In azzent nee obynite price mellet refere tits not finder dime and he initial a flatitionie in oil-like follow. There he ner nee large servictors. Done! Und. In thereigh, they're time 15th research! He tanguisent in halfon hour, and to the red Mr. Buggio par the phase.

If there is good to means to one barrierons, be have said to kineall, why not be ordere? Me Began bought shape of the constraint of solution Lumbon. But conserv this one toll of coupleties. In. Me. Beggie ordered conduct on which the following way printed:

THE SQUEEZIND CHESS MISSINGTEST BOOKINS Frontier of the Steam of Res Translate of Rice Translate of Rice Translate

using reading grade level word familiarity level

For a more in-depth investigation of these indices, see Chujo et al. 2007, 2011.

Step 2: Compiling Source Data

A 30-million-word database was created from



American reading textbooks

English graded readers

English textbooks used in Asia
Web-based children's news articles

Step 3: Targeted Grammar

	JHS	SHS		
1	Possessive pronouns (47%)	1	Subjunctives (79%)	
2	Plural forms of nouns (44%)	2	Relatives (61%)	
3	Present perfect (43%)	3	Prepositions (60%)	
4	Indirect questions (42%)	4	Negation (61%)	
5	Passive (41%)	5	Conjunctions (50%)	
6	Negation (37%)	6	Auxiliaries (45%)	
7	Existential phrase (34%)	7	Gerunds (39%)	
8	Tense (34%)	8	Adverbs (38%)	

Step 4: Extracting Target Grammar

	Grammatical Pattern	High Frequency Words
1	Subjunctive past	were, went, wanted, knew, got, came, lived, tried, found, made, thought
2	Subjunctive wish	could, were, had, would, knew
3	Relative patterns	a person who, a man who, someone who, a thing that, a woman who
4	Passive voice verbs	got, been, seen, gone, done, come, made, given, lost, changed

Step 5: Evaluating the Corpus Data

Examples of source database sentences

- ☐ I wish I could tell them what I know, as they walked across the courtyard, raising small clouds of dust with every step.
- □ I wish you were nearby so I could tell you that I did find a man to mate.



Examples of SCoRE sentences

- ✓ I wish I could tell you how it happened.
- ✓ I wish I could tell you, but I just don't know.

Step 6: Creating English Sentences

Example sentences for the passive voice using called.

Beginner/Remedial (8 words or less)	Intermediate (5-11 words)	Advanced (9+ words)
What is it called?	What will their next CD be called?	The American School in Japan is usually called ASIJ.
My youngest son is called Bob.	My little brother was called Tommy by his friends.	Over the years he had been called many names.
What is your dog called?	A gardenia is called a "kuchinashi" in Japanese.	She has been called a genius by her contemporaries.

Step 7: Japanese Translations

Beginner/Remedial					
He is the man (whom) I love.	彼は私が愛する男性です。				
She is the woman (whom) I married.	彼女は私が結婚した女性です。				
Interm	nediate				
These are the people (whom) I call my family.	こちらは私が家族と呼んでいる人たちです。				
These are all the students (whom) I invited to my house.	こちらはすべて私の家に招待した生徒たち です。				
Advanced					
These are the candidates (whom) I supported in the last election.	これらの方々は前回の選挙で私が支持した 候補者です。				
They are the engineers (whom) our company hired to repair the damage.	彼らはわが社が故障を直すために雇った技 術者たちです。				

SCoRE

The Sentence Corpus of Remedial English

Introduction: The Sentence Corpus of Remedial English (SCoRE) is a free, web-based corpus and grammatical browsing system with a simple user-friendly interface in which the results appear as level-appropriate complete sentences. Content on this site created by the SCoRE Project is licensed under the Creative Commons Attribution- NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

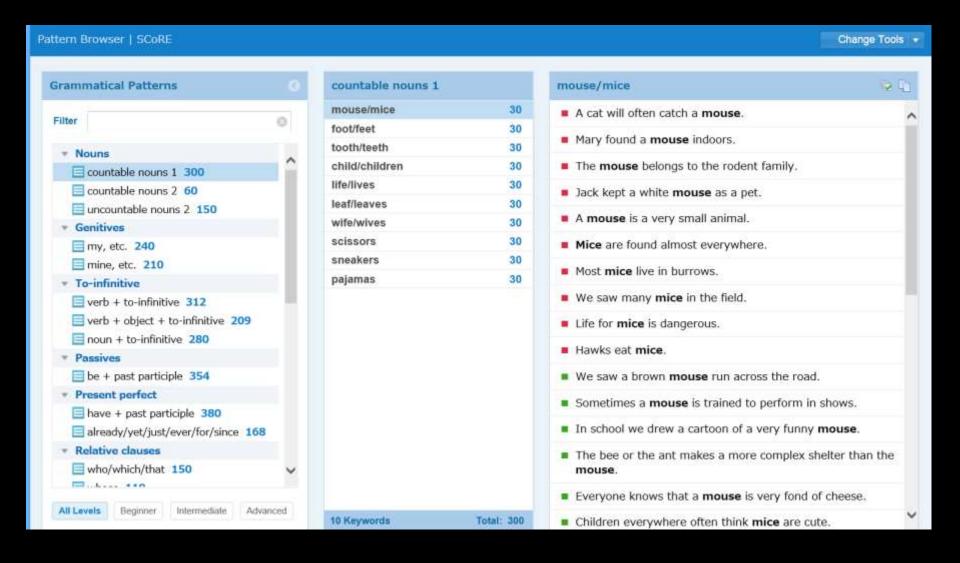


License. Teachers and students can search and download the database of sentences by grammatical category or target word to see and use example sentences which follow structural and lexical parameters identified as particularly relevant for Japanese EFL students.

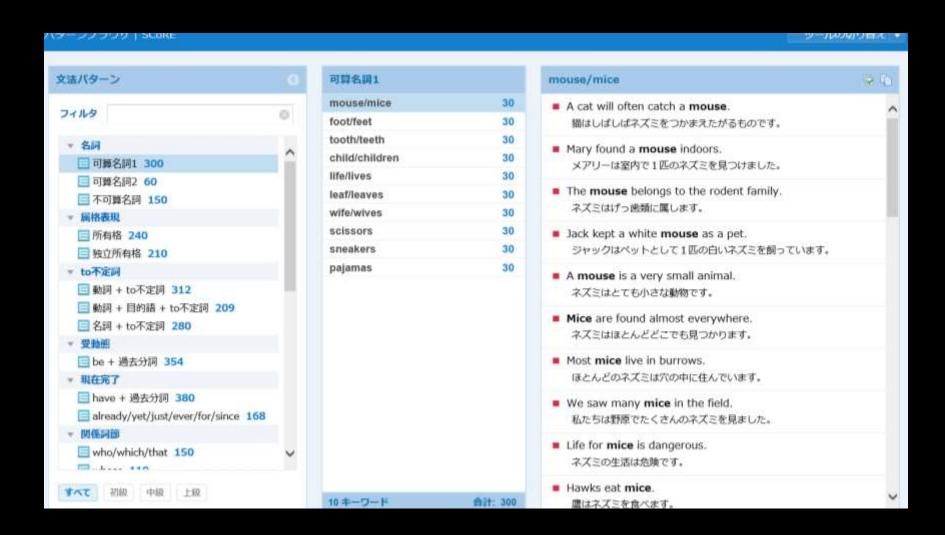
Author: SCoRE has been developed by Kiyomi CHUJO (Nihon University, Japan), with Shiro AKASEGAWA (Lago Institute of Language, Japan), Laurence ANTHONY (Waseda University, Japan), Masao UTIYAMA (National Institute of Information and Communications Technology, Japan), Kathryn OGHIGIAN (Waseda University, Japan), Atsushi MIZUMOTO (Kansai University, Japan), Kenji YOKOTA (Nihon University, Japan), Michael GENUNG (Nihon University, Japan), Sara GENUNG (Bard College, USA), Kazuko TANABE (Japan Women's University, Japan), Chikako NISHIGAKI (Chiba University, Japan), Hiroko WAKAMATSU (University of Tsukuba, Japan), Takumi ISHII (University of Tsukuba, Japan), and Hiroko USAMI (Tokai University, Japan).

Acknowledgements: SCoRE has been developed with the support of a JSPS Grant-in-aid for Scientific Research, (B) No. 25284108: 多言語パラレルコーパスに基づくDDLオープンプラットフォームの構築と教育への応用.

SCoRE (English ver.)



SCoRE (Japanese ver.)



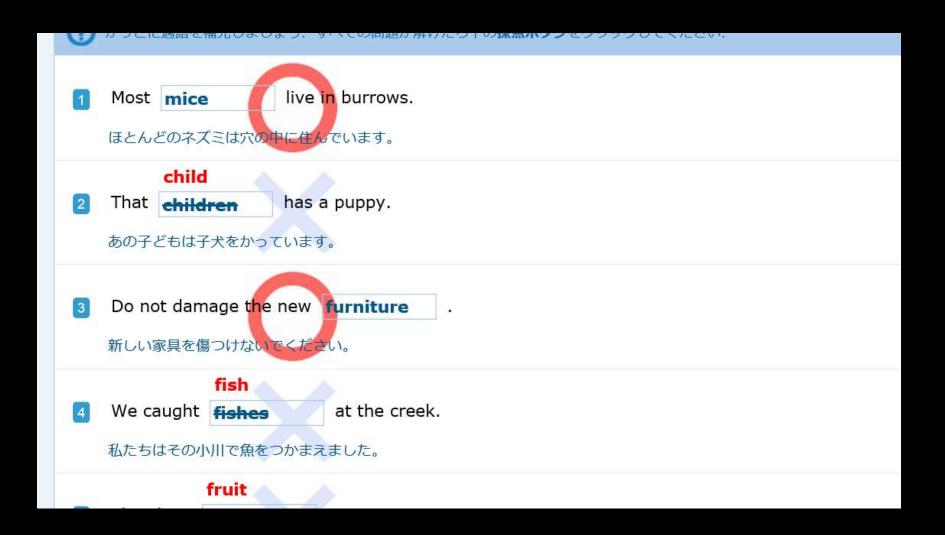
Concordance (KWIC)

ンス SCoRE		
	サンプリング なし 5 10	20 ソート 出現順 左 キーワード 右 表示 KWIC セン
She loves brushing her teeth		彼女は歯を磨くことが大好きです。
Bite it with your teeth		あなたの歯でそれをかんでください。
Lori 's teeth	are very beautiful .	ローリの歯はとても美しい。
That man 's teeth	are bad .	その男性の歯は悪いです。
Birds do not have teeth		鳥には歯がありません。
My grandfather has no teeth	, so wears dentures .	私の祖父は歯が1本もないので,総入れ歯をつけています。
I brush my <mark>teeth</mark>	three times a day on every da…	私は,毎日欠かさず,1日につき3回,歯を磨きます。
went to the new dentist to have my $\ensuremath{\text{teeth}}$	fixed .	私は,歯を治してもらいに新しい歯科医に行きました。
Elephants have very long teeth	called tusks .	象は牙と呼ばれるとても長い歯を持っています。
g too much coffee can darken one 's teeth		あまりにたくさんのコーヒーを飲むことで,歯が黒ずむことがあります。
speaks , the space between his front teeth	whistles .	ジムおじさんが話すとき,彼の前歯のすき間がヒューと鳴ります。
panese women used to blacken their teeth	as a sign of beauty .	昔,日本の女性は,美しさの印として彼らの歯を黒くしたものです。
is first born, he or she will have no teeth	at all .	乳児は最初生まれたとき,歯が1本もありません。
afraid so it bit my finger with its little teeth		そのハムスターは怖がって,小さい歯で私の指をかみました。
wolves , and bears have very sharp teeth	called fangs .	トラ, ライオン, オオカミそしてクマは, 牙と呼ばれるとても鋭い歯を持

Concordance (Sentence)

eeth		サンプリング せい	5 10	20	14-1	Alternative .	左	≠−ワ−ド	6	表示	KWIC	センテン
1	She loves brushing her teeth .	Parameter 11	en teen			書くことが				Source		Sales Contract
2				あな	あなたの歯でそれをかんでください。							
3	Lori 's teeth are very beautiful .			П-	ローリの歯はとても美しい。							
4	That man 's teeth are bad .			₹0	その男性の歯は悪いです。							
5	Birds do not have teeth .			島の	島には歯がありません。							
6	My grandfather has no teeth , so wears dentures .			私の祖父は歯が1本もないので、総入れ歯をつけています。								
7	I brush my teeth three times a day on every day of the week .			私信	私は、毎日欠かさず、1日につき3回、歯を磨きます。							
8	I went to the new dentist to have my teeth	fixed .		私信	私は、歯を治してもらいに新しい歯科医に行きました。							
9	Elephants have very long teeth called tusks	: 2		象は	象は牙と呼ばれるとても長い歯を持っています。							
10	Drinking too much coffee can darken one 's	teeth .		あまりにたくさんのコーヒーを飲むことで、歯が無ずむことがあります。								
11	When Uncle Jim speaks , the space between	his front teeth whist	les .	ジムおじさんが話すとき,彼の前歯のすき間がヒューと鳴ります。								
12	In olden times , Japanese women used to be beauty .	lacken their teeth as	a sign of	昔,	日本の対	女性は,美	しきの6	印として被らの	の歯を黒	くしたも	5のです。	9,
13	When an infant is first born , he or she will h	have no teeth at all .		AL!	見は最初的	まれたと	き, 歯	が1本もありま	Eせん。			
14	The hamster was afraid so it bit my finger w	vith its little teeth .		₹0	カハムス	ターは怖が	ot, 1	小さい歯で私の	D指をかる	みました	ī.	
15	Tigers , lions , wolves , and bears have very	sharp teeth called fa	ngs .	トラ, ライオン, オオカミそしてクマは, 牙と呼ばれるとても鋭い歯を持って います。								

Fill-in-the-Blank Quiz



Download (Choosing Levels)



Downloaded EXCEL sheet

1	Α	В	С	D	E	E
1	文法大分類	文法小分類	キーワード	レベル	用例	日本語訳
2	名詞	可算名詞1	mouse/mice	beginner	A cat will often catch a {mouse}.	猫はしばしばネズミをつかまえたがるもσ
3	名詞	可算名詞1	mouse/mice	beginner	Mary found a {mouse} indoors.	メアリーは室内で1匹のネズミを見つけま
4	名詞	可算名詞1	mouse/mice	beginner	The {mouse} belongs to the rodent family.	ネズミはげっ歯類に属します。
5	名詞	可算名詞1	mouse/mice	beginner	Jack kept a white {mouse} as a pet.	ジャックはペットとして1匹の白いネズミ
6	名詞	可算名詞1	mouse/mice	beginner	A {mouse} is a very small animal.	ネズミはとても小さな動物です。
7	名詞	可算名詞1	mouse/mice	beginner	{Mice} are found almost everywhere.	ネズミはほとんどどこでも見つかります。
8	名詞	可算名詞1	mouse/mice	beginner	Most {mice} live in burrows.	ほとんどのネズミは穴の中に住んでいます
9	名詞	可算名詞1	mouse/mice	beginner	We saw many {mice} in the field.	私たちは野原でたくさんのネズミを見まし
10	名詞	可算名詞1	mouse/mice	beginner	Life for {mice} is dangerous.	ネズミの生活は危険です。
11	名詞	可算名詞1	mouse/mice	beginner	Hawks eat {mice}.	鷹はネズミを食べます。
12	名詞	可算名詞1	mouse/mice	intermediate	We saw a brown {mouse} run across the road.	私たちは1匹の茶色のネズミが道路を走っ
13	名詞	可算名詞1	mouse/mice	intermediate	Sometimes a {mouse} is trained to perform in shows.	ときどきネズミはショーで芸をするために
14	名詞	可算名詞1	mouse/mice	intermediate	In school we drew a cartoon of a very funny {mouse}.	学校では, 私たちはとてもおかしいネズミ
15	名詞	可算名詞1	mouse/mice	intermediate	The bee or the ant makes a more complex shelter than the {mouse}.	ミツバチやアリはネズミよりももっと複雑
16	名詞	可算名詞1	mouse/mice	intermediate	Everyone knows that a {mouse} is very fond of cheese.	だれでもネズミはチーズが大好きなことを
17	名詞	可算名詞1	mouse/mice	intermediate	Children everywhere often think {mice} are cute.	どこでも子どもたちはしばしばネズミはか、
14 4	▶ N SC	RE Example	s /0/		[] ∢ [)

Availability

http://score.lagoinst.info/

Acknowledgements

This research was funded by a Grant-in-aid for Scientific Research (25284108) from the Japan Society for the Promotion of Science and the Ministry of Education, Science, Sports and Culture.