東京外国語大学 Tokyo University of Foreign Studies

Corpus-based Research on Criterial Features for the CEFR levels

YUKIO TONO
Tokyo University of Foreign Studies

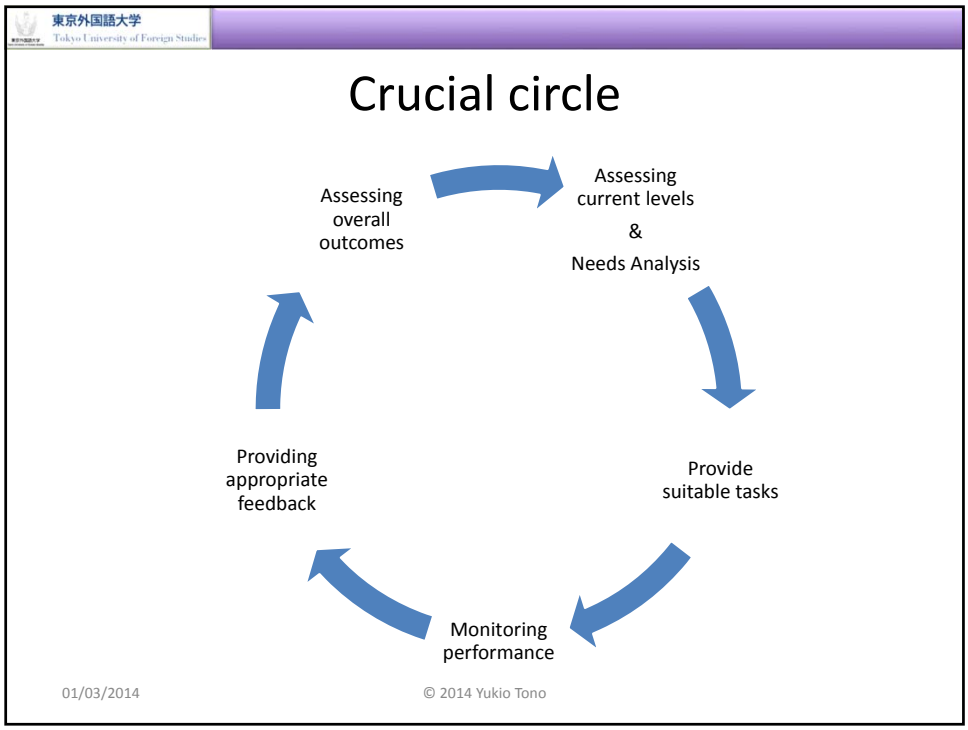
"Language Education and Computer Science for Second Language Acquisition, e-Learning and Learner Corpora"
1 March, 2014, TUFS

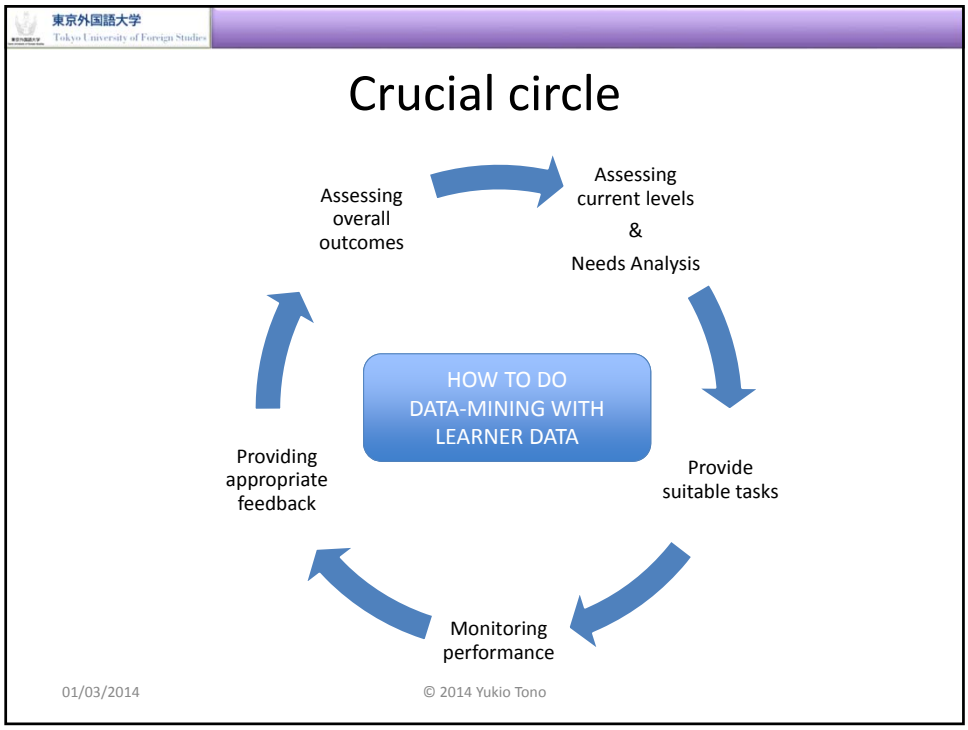
東京外国語大学 Tokyo University of Foreign Studies

CALL environment

- More and more opportunities of using CALL because of the network-friendly environment at school
- Accessibility of hand-held devices (mobiles, tablet PCs) → provide ubiquitous environment for learning
- More and more contents have been created for general audience (e.g. TED Talk).
- But CALL materials are still mostly based on the reuse of previously published materials in ELT.

© 2014 Yukio Tono


東京外国語大学
Tokyo University of Foreign Studies
Crucial circle
Assessing current levels & Needs Analysis
Provide suitable tasks
Monitoring performance
Providing appropriate feedback
Assessing overall outcomes
01/03/2014
© 2014 Yukio Tono


東京外国語大学
Tokyo University of Foreign Studies
Crucial circle
Assessing current levels & Needs Analysis
Provide suitable tasks
Monitoring performance
Providing appropriate feedback
Assessing overall outcomes
HOW TO DO DATA-MINING WITH LEARNER DATA
01/03/2014
© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Crucial circle

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Assessing current levels

- Standardized tests → If the course is designed to prepare for TOEFL, then it's OK, but usually test scores provide very limited range of information.
- Needs Analysis:
  - Target Situation Analysis: what is the ideal situation?
  - Deficiency (Present Situation) Analysis: what is the present situation and what is lacking? How can we fill the gap?
- Standardized tests often do not provide answers to that kind of question.

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

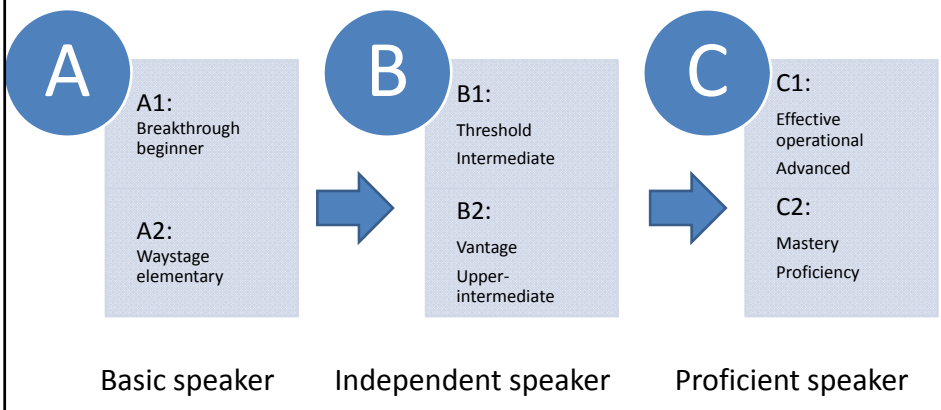
CEFR: A Better Benchmark

- CEFR: Common European Framework of Reference for Languages
- CEFR CAN-DO descriptors
  → 5 skills (SP-I, SP-P, L, R, W) & 6 levels (A1, A2, B1, B2, C1, C2) accompanied by detailed CAN-DO statements
  → self-assessment based on CAN-DO
- Self-assessment grid is available free and it has been widely used in Europe for developing & assessing foreign language syllabuses and course contents.

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

CEFR-levels



© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Table 2. *Common Reference Levels: self-assessment grid*

		A1	A2	B1
UNDERSTANDING	**Listening**	I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.
	Reading	I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.	I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.
SPEAKING	**Spoken Interaction**	I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics.	I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.	I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).
	Spoken Production	I can use simple phrases and sentences to describe where I live and people I know.	I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.
WRITING	**Writing**	I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.	I can write short, simple notes and messages relating to matters in areas of immediate need. I can write a very simple personal letter, for example thanking someone for something.	I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Mining CAN-DOs

- Self-assessment using CEFR CAN-DO descriptors → Each student provides answers to whether he/she can do well with the task described in, say, 100 CAN-DO descriptors.
- Each student has the scores such as follows:
  - Student 1: [1, 2, 5, 8, 13, 19, … 56]
  - Student 2: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, …, 89]
  - Student 3: [1, 2, 7, 11, …, 33]
- This is exactly the same as "market basket transaction."

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Market basket transaction

- Point of Sale (POS) data:
- Type of data: market basket transaction:

TID	ITEMS
1	{bread, milk, ham, apple}
2	{bread, ham, noodle, egg}
3	{sausage, beer, cheeze, yogurt}
4	{bread, milk, butter, egg}

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Association Analysis

- X ⇒ Y (association rule):
- {onions, potatoes} ⇒ {burger}
  - if a customer buys onions and potatoes together, he or she is likely to also buy hamburger meat.
- Association rule learning is a method for discovering interesting relations between variables in large databases. (Wikipedia)
- There is a R implementation of association analysis ("arules" package)

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

R: arules (mock example)

```
> library(arules)
> # find association rules with default settings
> rules <- apriori(cefr.raw)
> inspect(rules)
  lhs              rhs         support   confidence lift
1 {}            => {A1-all}    0.9504771 0.9504771  1.0000000
2 {A1-W}        => {A1-all}    0.3585825 0.9157895  1.5435051
3 {B1-SI}       => {A2-all}    0.2449341 0.9815385  2.1126798
4 {A2-SP}       => {A2-W}      0.1930940 0.9042553  0.9513700
```

You could have better understanding of what learners can or cannot do if they can do well with particular can-do's.

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

CAN-DO based needs analysis

	A1	A2	B1	B2	C1	C2
SP-I						
SP-P						
W						
L						
R						

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

CAN-DO based needs analysis

	A1	A2	B1	B2	C1	C2
SP-I						
SP-P						
W						
L						
R						

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

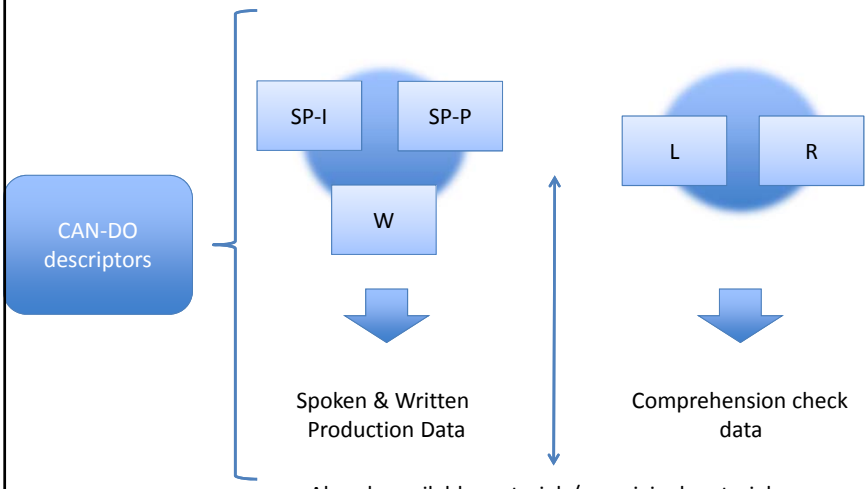
CAN-DO based needs analysis

	A1	A2	B1	B2	C1	C2
SP-I						
SP-P						
W						
L						
R						

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

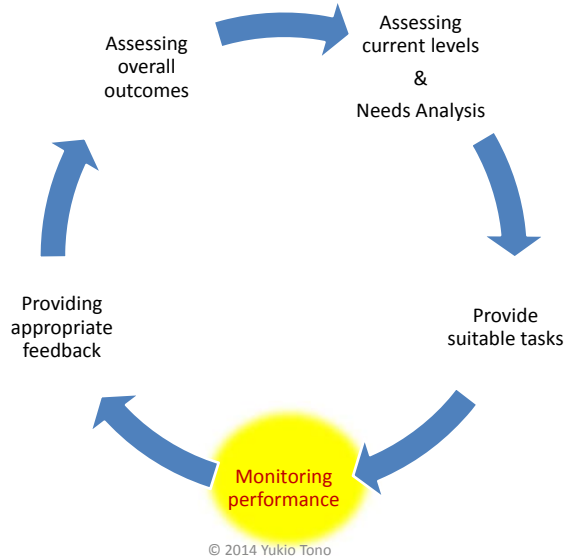
Linking CEFR to the course contents



© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Crucial circle



© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Mining production data

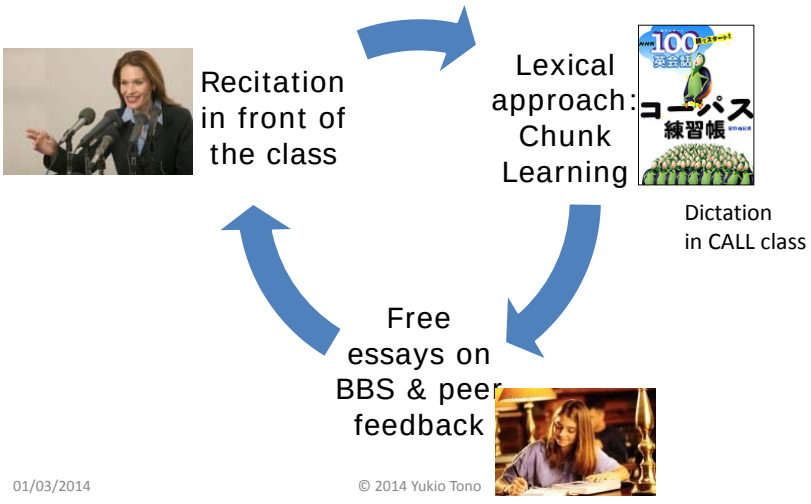
- Creating corpus of learner output
  → learner corpora
- Not in traditional sense, though

- Your own learners' corpus
- "Local learner corpora" (Seidlehofer 2002)
  - The learner data comes from your own students and will be used to provide useful feedback to improve the course.
  - Action research oriented perspectives

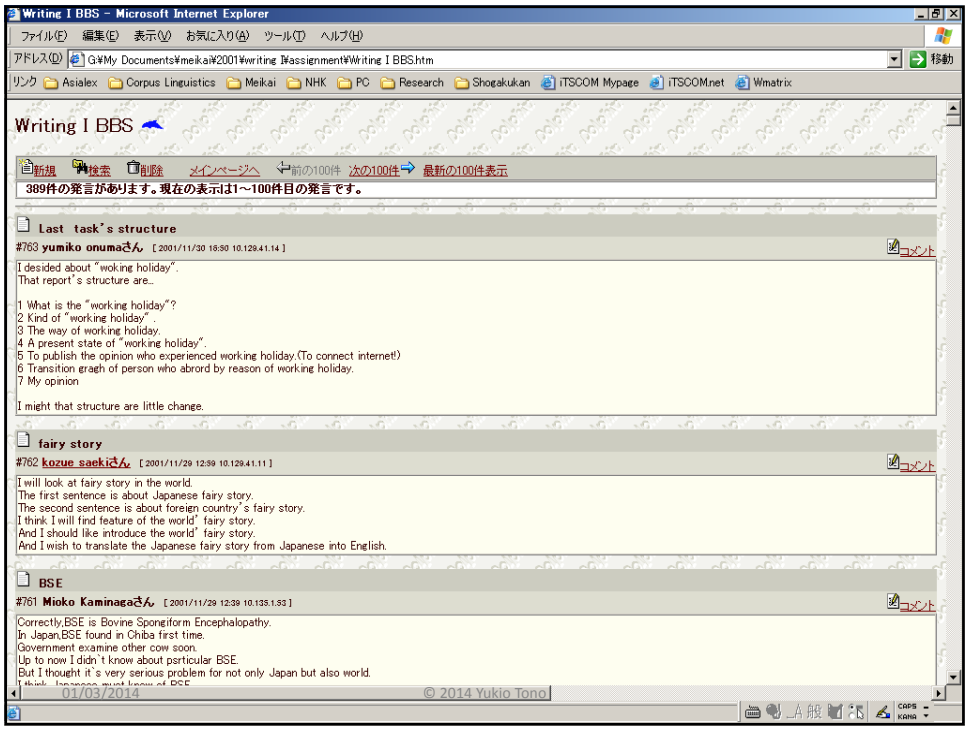
© 2014 Yukio Tono

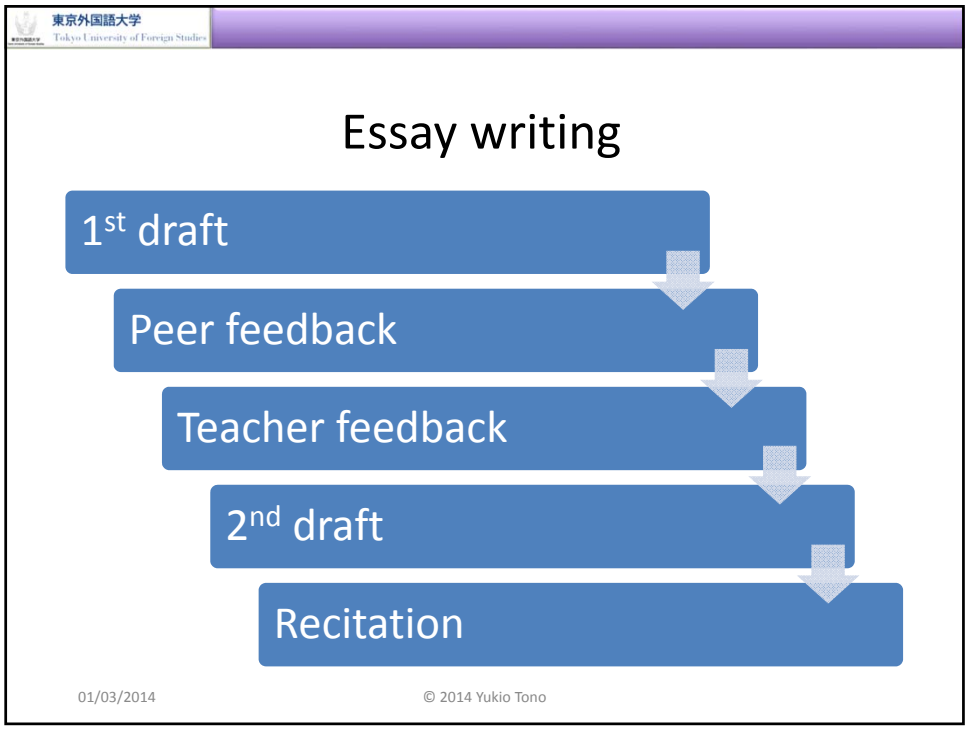
東京外国語大学 Tokyo University of Foreign Studies

My previous experience in L2 writing course



© 2014 Yukio Tono


Writing I BBS - Microsoft Internet Explorer
Writing I BBS
389件の発言があります。現在の表示は1～100件目の発言です。
Last task's structure
#763 yumiko onumaさん
I desided about "woking holiday".
That report's structure are...
1 What is the "working holiday"?
2 Kind of "working holiday".
3 The way of working holiday.
4 A present state of "working holiday".
5 To publish the opinion who experienced working holiday.(To connect internet!)
6 Transition gragh of person who abrord by reason of working holiday.
7 My opinion
I might that structure are little change.
fairy story
#762 kozue saekiさん
I will look at fairy story in the world.
The first sentence is about Japanese fairy story.
The second sentence is about foreign country's fairy story.
I think I will find feature of the world' fairy story.
And I should like introduce the world' fairy story.
And I wish to translate the Japanese fairy story from Japanese into English.
BSE
#761 Mioko Kaminagaさん
Correctly,BSE is Bovine Spongiform Encephalopathy.
In Japan,BSE found in Chiba first time.
Government examine other cow soon.
Up to now I didn't know about psrticular BSE.
But I thought it's very serious problem for not only Japan but also world.
01/03/2014
© 2014 Yukio Tono


東京外国語大学
Tokyo University of Foreign Studies
Essay writing
1st draft
Peer feedback
Teacher feedback
2nd draft
Recitation
01/03/2014
© 2014 Yukio Tono

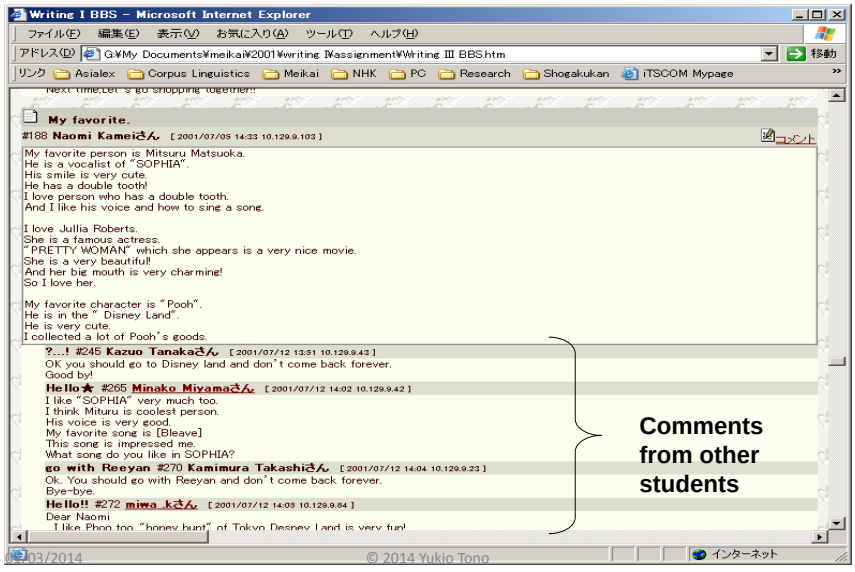
東京外国語大学 Tokyo University of Foreign Studies

Crucial circle

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Write a free composition → peer feedback about contents

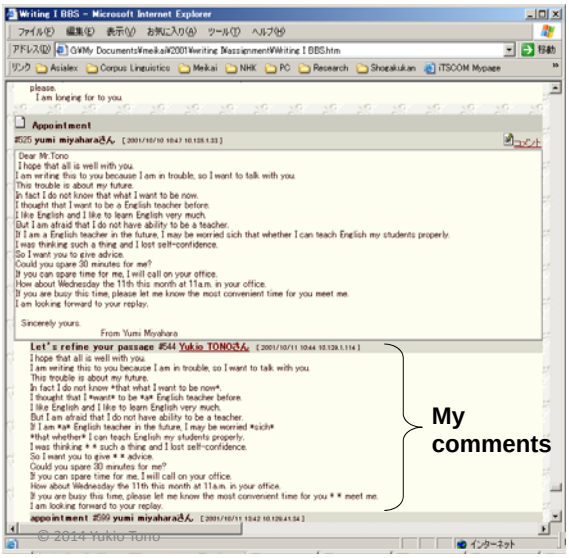


© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Comments on grammar & vocabulary from teachers

- Coded feedback
- Students were asked to revise & re-submit the draft.
- Students were told to recite the essays.
- 10 students every class. Every once a month, video recitation.



© 2014 Yukio Tono

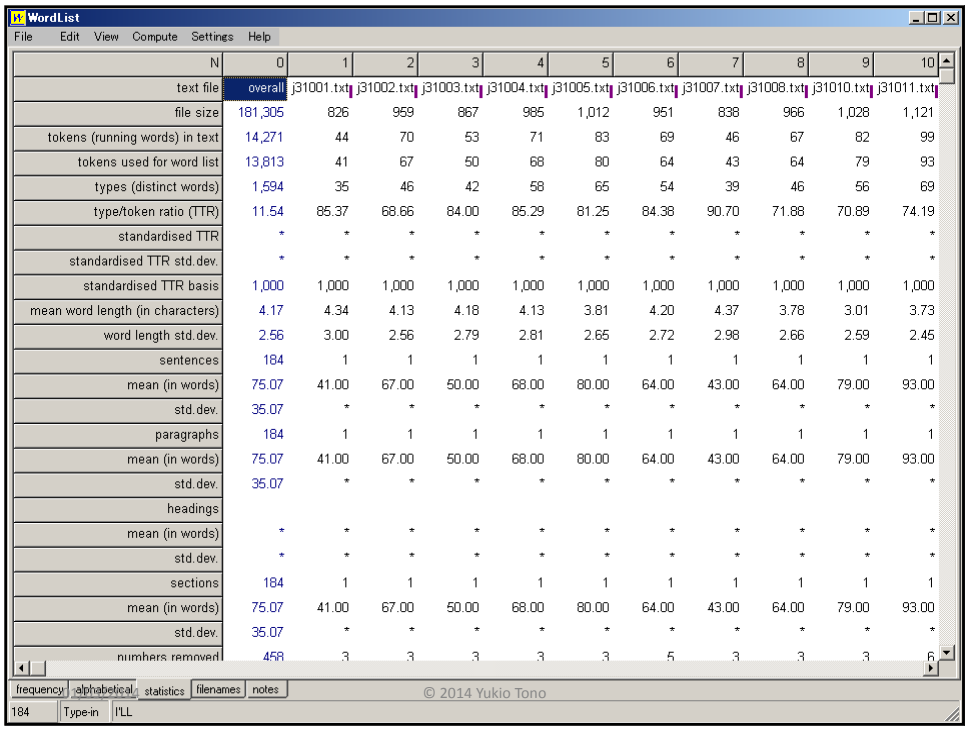
東京外国語大学 Tokyo University of Foreign Studies

Evaluation of the essays

- More objective measures for writing:
  - Total number of words per essay
  - Lexical diversity (type-token ratio; Guiraud)
  - Complexity (Average sentence length)
- Return the evaluation sheet with the above statistics along with the class average stats

→ Students could appreciate how good their writing is compared to the whole class in terms of the above criteria.

© 2014 Yukio Tono



WordList

File Edit View Compute Settings Help

N	0	1	2	3	4	5	6	7	8	9	10
text file	overall	j31001.txt	j31002.txt	j31003.txt	j31004.txt	j31005.txt	j31006.txt	j31007.txt	j31008.txt	j31010.txt	j31011.txt
file size	181,305	826	959	867	985	1,012	951	838	966	1,028	1,121
tokens (running words) in text	14,271	44	70	53	71	83	69	46	67	82	99
tokens used for word list	13,813	41	67	50	68	80	64	43	64	79	93
types (distinct words)	1,594	35	46	42	58	65	54	39	46	56	69
type/token ratio (TTR)	11.54	85.37	68.66	84.00	85.29	81.25	84.38	90.70	71.88	70.89	74.19
standardised TTR	*	*	*	*	*	*	*	*	*	*	*
standardised TTR std.dev.	*	*	*	*	*	*	*	*	*	*	*
standardised TTR basis	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
mean word length (in characters)	4.17	4.34	4.13	4.18	4.13	3.81	4.20	4.37	3.78	3.01	3.73
word length std.dev.	2.56	3.00	2.56	2.79	2.81	2.65	2.72	2.98	2.66	2.59	2.45
sentences	184	1	1	1	1	1	1	1	1	1	1
mean (in words)	75.07	41.00	67.00	50.00	68.00	80.00	64.00	43.00	64.00	79.00	93.00
std.dev.	35.07	*	*	*	*	*	*	*	*	*	*
paragraphs	184	1	1	1	1	1	1	1	1	1	1
mean (in words)	75.07	41.00	67.00	50.00	68.00	80.00	64.00	43.00	64.00	79.00	93.00
std.dev.	35.07	*	*	*	*	*	*	*	*	*	*
headings											
mean (in words)	*	*	*	*	*	*	*	*	*	*	*
std.dev.	*	*	*	*	*	*	*	*	*	*	*
sections	184	1	1	1	1	1	1	1	1	1	1
mean (in words)	75.07	41.00	67.00	50.00	68.00	80.00	64.00	43.00	64.00	79.00	93.00
std.dev.	35.07	*	*	*	*	*	*	*	*	*	*
numbers removed	458	3	3	3	3	3	5	3	3	3	6

frequency alphabetical statistics filenames notes

© 2014 Yukio Tono

184 Type-in I'LL

東京外国語大学 Tokyo University of Foreign Studies

Mining your LC

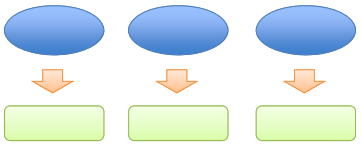
- As the learner data will accumulate, you might want to exploit the data more fully. Here are some of the studies I conducted on my learner corpora.
- Learning order of verb co-occurrence patterns
  - Parsed data + HCFA
- Automated error annotation
  - Edit distance
- Feature extraction for CEFR levels
  - Main topic for the main conference
  - Machine learning

© 2014 Yukio Tono

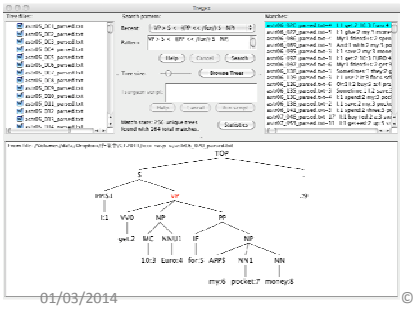
東京外国語大学 Tokyo University of Foreign Studies

Tono (2011): verb co-occurrence patterns

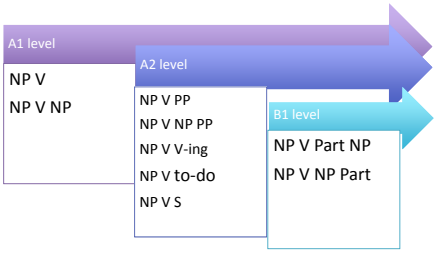
CEFR-based subcorpora



Parsed with Stanford Parser



Extract verb patterns using Tregex



Tono (2011)

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

RASP (G-default output)

- (|My:1_APP$| |favorite:2_JJ| |food:3_NN1| |be+s:4_VBZ| |KimChi:5_NP1| |.:6_.|) 1 ; (-7.664)
- gr-list: 1
- (|ncsubj| |be+s:4_VBZ| |food:3_NN1| _)
- (|xcomp| _ |be+s:4_VBZ| |KimChi:5_NP1|)
- (|det| |food:3_NN1| |My:1_APP$|)
- (|ncmod| _ |food:3_NN1| |favorite:2_JJ|)

- RASP: an open-source statistical parser developed by Ted Briscoe, John Carroll, et al. (Briscoe, Carroll, & 2006)

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

RASP (UPENN format)

- (|My:1_APP$| |favorite:2_JJ| |food:3_NN1| |be+s:4_VBZ| |KimChi:5_NP1| |.:6_.|) 1 ; (-7.664)
- upenn: 1
- (TOP
- (S (NP (APP$ My:1) (JJ favorite:2) (NN1 food:3))
- (VP (VBZ be+s:4) (NP1 KimChi:5)))
- (. .:6))

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

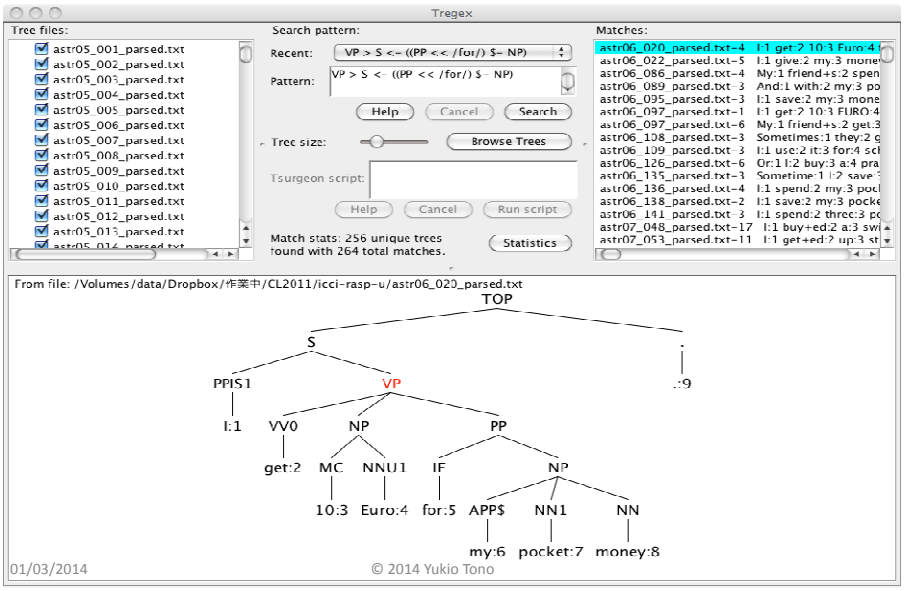
Tregex

Co-occurrence frames	Tregex syntax
NP-V	VP > S <- (/VV.*/ !,, /V[VMHB].*/)
NP-V (reciprocal Subj.)	VP > S <- (/VV.*/ !,, /V[VMHB].*/)
NP-V-PP	VP > S <- (PP !,, NP)
NP-V-NP	VP > S <- (NP !< PP)
NP-V-Part-NP	RP > (V $++ NP)
NP-V-NP-Part	VP << (RP $- /(PP[IH]..|NP)/)
NP-V-NP-PP	VP > S <- (PP $- NP)
NP-V-NP-PP (P=for)	VP > S <- ((PP << /for/) $- NP)
NP-V-V (+ing)	VP > S < (/VV.*/ . VVG)
NP-V-Vpinfinitaival (Subj Control)	VP > S < (/VV.*/ . (TO . VV0))
NP-V-S	VP > S <- (S , (/VV.*/ << /think/))

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Tregex gui



© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Data analysis

1. Using Tregex, 98,675 sentences were processed for verb patterns.
2. Creation of matrix (verb pattern X region X CEFR　X　freq）
3. A hierarchical configural frequency analysis (von Eye 1990; Gries 2004; 2009).
   R script provided by S. Th. Gries (HCFA 3.2).

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

A1 x v-pattern

VPATTERN	CEFR	Freq	Exp	Cont.chisq	Obs-exp	*P*Holm	Dec	Q
v-n	A1	12366	10993.0	171.4685	>	2.71E-43	***	0.023
v-s	A1	66	688.553	562.8798	<	9.75E-205	***	0.009
v-to-v	A1	1636	1959.21	53.3221	<	4.67E-13	***	0.005
intrans	A1	131	88.2761	20.6775	>	0.0002403	***	0.001
v-n-part	A1	4	73.9472	66.1636	<	3.42E-25	***	0.001
v-n-pp	A1	614	718.2348	15.1272	<	0.0006127	***	0.001
v-n-pp_for	A1	150	215.7007	20.0119	<	2.98E-05	***	0.001
v-part-n	A1	54	122.563	38.3548	<	7.76E-11	***	0.001
v-pp	A1	2933	3025.43	2.8244	<	0.3896257	ns	0.001
v-v-ing	A1	269	338.00	14.0887	<	0.0009505	***	0.001

- For A1 level, only [v-n] and [intrans] were "types" for this level. The rest were antitype.

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

A2 x v-pattern

VPATTERN	CEFR	Freq	Exp	Cont.chisq	Obs-exp	*P*Holm	Dec	Q
v-pp	A2	4718	4406.0874	22.0807	>	1.88E-05	***	0.005
v-n	A2	16166	16009.7033	1.5259	>	0.56770530	ns	0.003
v-to-v	A2	2669	2853.2993	11.9042	<	0.00273645	**	0.003
intrans	A2	89	128.5606	12.1736	<	0.00209686	**	0.001
v-n-part	A2	44	107.6928	37.6699	<	7.85E-11	***	0.001
v-part-n	A2	122	178.4942	17.8807	<	9.58E-05	***	0.001
v-s	A2	954	1002.7724	2.3722	<	0.49186688	ns	0.001
v-v-ing	A2	446	492.2565	4.3466	<	0.19890446	ns	0.001
v-n-pp	A2	1040	1045.9986	0.0344	<	0.86785746	ns	0
v-n-pp_for	A2	291	314.1349	1.7038	<	0.49590241	ns	0

- [v-pp] was found to be a significant type for A2 level.
- [n-part] and [part-n] were found to be "antitypes".

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

結果: B1 x V-pattern

VPATTERN	CEFR	Freq	Exp	Cont.chisq	Obs-exp	*P*Holm	Dec	Q
v-n	B1	12915	14138.415	105.8638	<	6.24E-30	***	0.021
v-s	B1	1423	885.5638	326.1625	>	2.70E-61	***	0.008
v-to-v	B1	2963	2519.7926	77.9559	>	3.63E-17	***	0.006
v-pp	B1	3630	3891.0837	17.5182	<	0.0001421	***	0.004
v-n-pp	B1	1039	923.7375	14.3823	>	0.0015193	**	0.002
v-v-ing	B1	551	434.7193	31.1033	>	1.06E-06	***	0.002
v-n-part	B1	199	95.1051	113.4971	>	3.47E-19	***	0.001
v-n-pp_for	B1	370	277.4174	30.8976	>	1.56E-06	***	0.001
v-part-n	B1	248	157.631	51.8081	>	4.94E-10	***	0.001
intrans	B1	99	113.5338	1.8605	<	0.5497820	ns	0

- [v-n] はB1　では antitype となった。低レベルの指標だからである.
- [v-s]はB1レベルでより有意味な指標として判定された.
- [n-part] と [part-n] はここで 有効な指標として判定されたので、A2 から移動するべき。

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Revision of williams (2007) based on our Japanese EFL learners' data

Co-occurrence frames	Criterial for CEFR Level:
NP-V	A2 → A1
NP-V-PP	A2
NP-V-NP	A2 → A1
NP-V-Part-NP	A2 → B1
NP-V-NP-Part	A2 → B1
NP-V-NP-PP	A2
NP-V-NP-PP (P=for)	A2
NP-V-V (+ing)	A2 (strong predictor for B1)
NP-V-Vpinfinitaival (Subj Control)	A2 (strong predictor for B1)
NP-V-S	A2 (strong predictor for B1)

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Tono & Mochizuki (2009): Semi-automatic error annotation

Original writings ⟺ Proofread versions

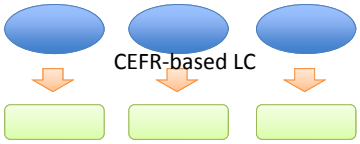
Sentence X: … A …	⟺	Sentence X': … …	⇒ Sentence X: … <add>A</add> …
Sentence X: … …	⟺	Sentence X': … A …	⇒ Sentence X: … <oms>A</oms> …
Sentence X: … A …	⟺	Sentence X': … B …	⇒ Sentence X: … <msf crr="B">A</msf> …

Edit distance

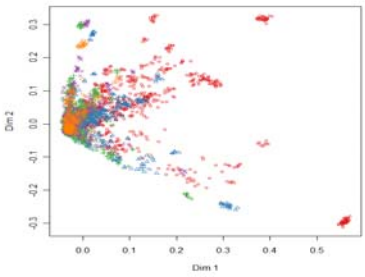
© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Tono (2011): Machine learning



Omission
Addition × POS
Misformation



Random Forest (Machine learning)

Contribution of different types of errors to distinguish CEFR levels
↓
Feature gravity

© 2014 Yukio Tono

東京外国語大学 Tokyo University of Foreign Studies

Crucial circle

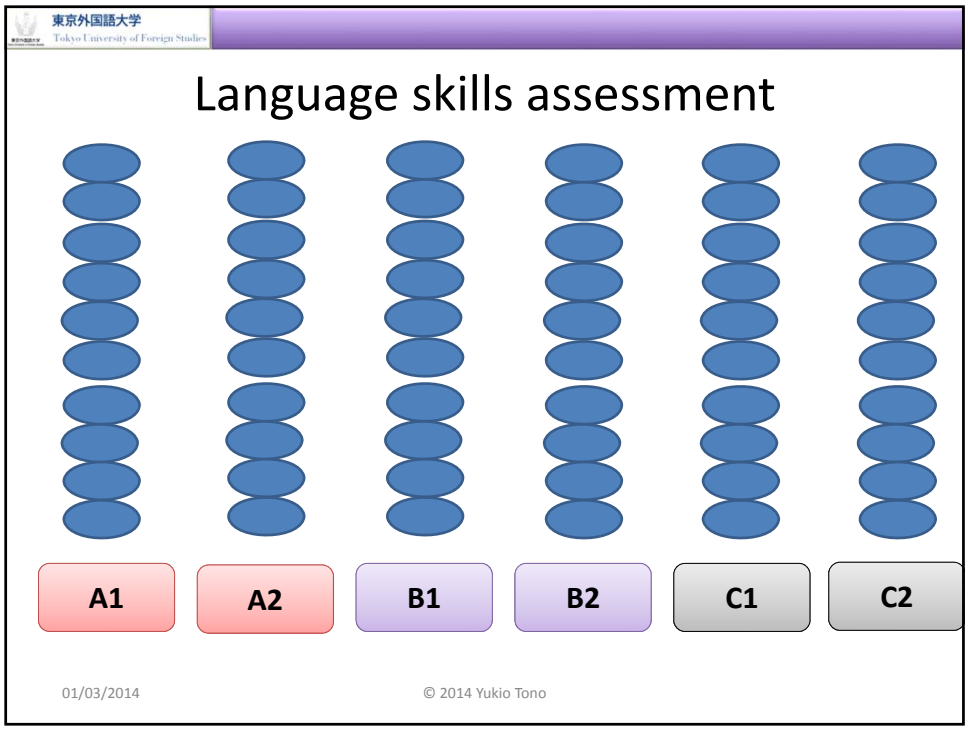
© 2014 Yukio Tono

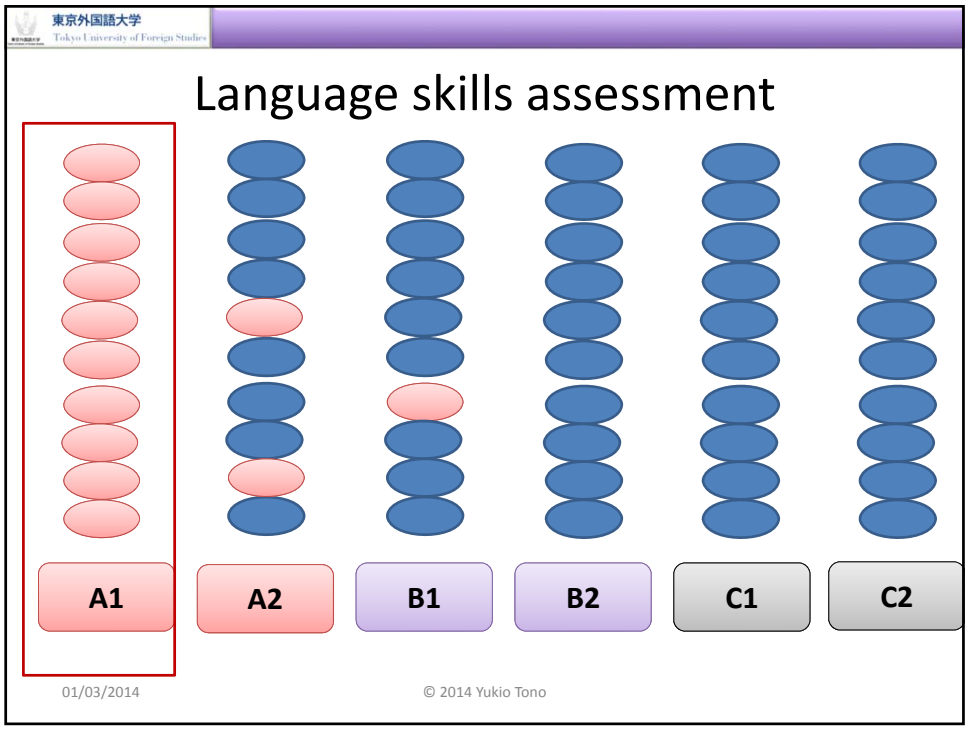
東京外国語大学 Tokyo University of Foreign Studies

Performance assessment

- More formative assessment will be used.
- 4 skills will be integrated.
- Not only performance but also self-assessment using can-do questionnaires will be extensively used.
- The future assessment system will combine day-to-day performance in classroom as well as occasional can-do self-assessment with portfolios.
- The overall image of the skills assessment will be not based on discrete-point tests, but more like the combinations of different can-do tasks.

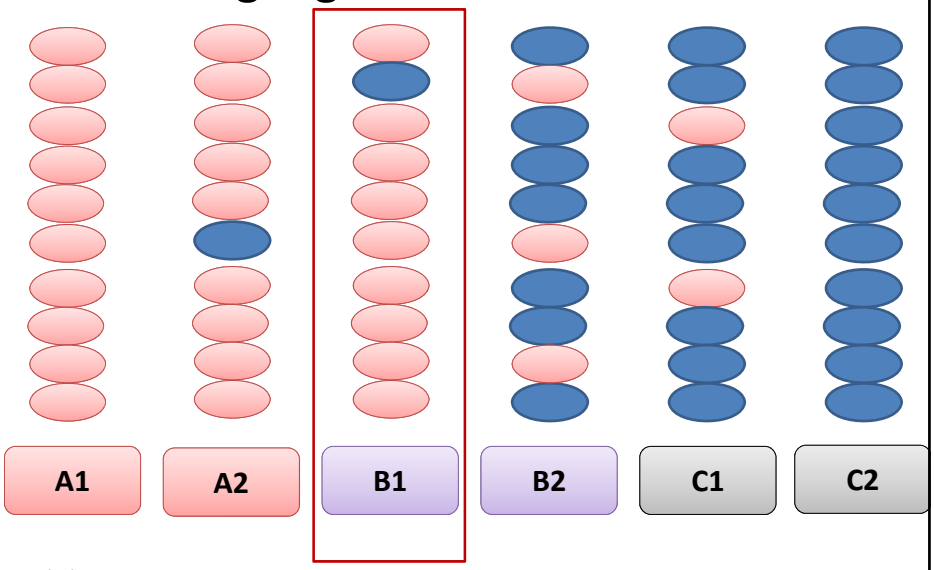
© 2014 Yukio Tono


東京外国語大学
Tokyo University of Foreign Studies
Language skills assessment
A1
A2
B1
B2
C1
C2
01/03/2014
© 2014 Yukio Tono


東京外国語大学
Tokyo University of Foreign Studies
Language skills assessment
A1
A2
B1
B2
C1
C2
01/03/2014
© 2014 Yukio Tono

東京外国語大学
Tokyo University of Foreign Studies

Language skills assessment



© 2014 Yukio Tono

東京外国語大学
Tokyo University of Foreign Studies

Thanks!

y.tono@tufs.ac.jp