

# ปัญหาการเรียนภาษาต่างประเทศ

- ปัญหาของการเรียนภาษาต่างประเทศคืออะไร
- มีสาเหตุมาจากอะไร
- แนวทางแก้ไขคืออะไร
- การใช้ corpus จะช่วยแก้ปัญหาดังกล่าวอย่างไร

# Corpus กับการเรียนรู้การสอนภาษา

- ช่วยด้านการจัดทำหลักสูตรการสอนภาษา (pedagogy)
- มีมาตั้งแต่ 1920s เช่น งานของ Thronrdike (1921)
- เกิดแนวทางการคัดเลือกคำศัพท์ที่เหมาะสมสำหรับผู้เรียน (vocabulary control)
- จาก Carnegie Report ปี 1936 (ผลจากการประชุมทางวิชาการด้านการสอนภาษาอังกฤษในปี ค.ศ. 1934 และ 1935 ) เสนอประเด็นว่า ควรให้ความสำคัญกับเรื่องของความถี่ของคำ เน้นให้ความรู้เกี่ยวกับคำที่มีความถี่ในการปรากฏใช้สูงเป็นหลัก

# ประโยชน์ด้านการเรียนการสอน ภาษา

- Michael West (1953) จัดทำ General Service List of English Words ซึ่งถือว่าเป็นฉบับปรับปรุงของ Carnegie Report ในปี ค.ศ. 1936
- West ยังสนใจนับความแตกต่างทางความหมายของคำแต่ละคำด้วย โดยได้แจกแจงความถี่ของความหมายต่างๆของคำจำนวน 2,000 คำ ซึ่งคัดมาจากรายการคำที่มีความถี่สูงสุด 2,000 คำแรก
- ช่วงปลายทศวรรษ 1940 เป็นต้นมา ความสนใจในทฤษฎีการสอนภาษา ได้เปลี่ยนจากการเน้นความรู้เรื่องคำศัพท์ (vocabulary) มาเน้นที่ โครงสร้างภาษา (structural syllabus) ทั้งนี้เป็นผลสืบเนื่องจาก แนวคิดของไวยากรณ์โครงสร้าง

# ประโยชน์ด้านการเรียนการสอน ภาษา

- ยุค Chomsky ความสนใจเรื่องข้อมูลภาษาจริงลดลง
- อ้างอิงข้อจำกัดเรื่องข้อมูล ไม่ representative มีเฉพาะภาษาเขียน
- ต่อมา แนวคิด functional-notional เข้ามามีบทบาททำให้เกิดหลักสูตรการสอนภาษาแบบ functional-notional syllabus ซึ่งให้ความสำคัญกับ situation มากขึ้น เกิดเป็นแนวทางแบบการสอนภาษาเพื่อการสื่อสาร (communicative approach)
- เหมาะกับการสอนเพื่อให้ใช้งานได้จริง
- สนใจเทคนิคการสอนและอื่นๆ ไม่ได้สนใจเฉพาะเรื่องภาษา

# ประโยชน์ด้านการเรียนการสอน ภาษา

- ความหมายเป็นสิ่งที่ได้จากการใช้ในสถานการณ์นั้นๆ ไม่ใช่มาจากการผนวกความหมายย่อยๆ ของคำแต่ละคำ (composition)
- อาจารย์ผู้สอนภาษาได้เปลี่ยนบทบาทจากการเป็น source of knowledge มาเป็น organizer of tasks or opportunities for learning
- Sinclair (1998) : อาจารย์ที่สอนตามแนว communicative language teaching ไม่ได้สนใจอธิบายภาษาที่เกิดขึ้นในปัจจุบัน กลับไปเน้นในเรื่องเทคนิคต่างๆ ในการเรียนการสอน ทำเสมือนว่าความรู้ความเข้าใจในภาษาอังกฤษนั้นได้ถูกค้นพบสมบูรณ์แล้ว

# ประโยชน์ด้านการเรียนการสอน ภาษา

- ปัจจุบัน เริ่มให้ความสนใจกับการใช้ corpus ในการเรียนการสอน
- corpus ให้ข้อมูลความถี่ เห็นรูปแบบการใช้แบบต่างๆ
- ให้ข้อมูลเรื่องคำปรากฏรวม collocation ซึ่งจำเป็นสำหรับผู้เรียน
- ให้เข้าใจความหมายและการใช้ในลักษณะต่างๆ ของคำได้ดีมากขึ้น
- เห็นว่าสิ่งที่เข้าใจและสรุปกันเป็นกฎไวยากรณ์นั้นบางครั้งก็ไม่มีปรากฏใช้จริง หรือสิ่งที่พบจริง ก็อาจจะไม่ได้มีการสรุปเป็นกฎไวยากรณ์เอาไว้

# Corpus ling and lang teaching

- ◆ Fligelstone (1993) describes three aims of corpus-based linguistics in teaching:
- ◆ teaching about (the principles and theory behind the use of corpora),
- ◆ teaching to exploit (the practical, methodological aspects of corpus-based work), and
- ◆ exploiting to teach (using corpora to derive or drive teaching materials).
- ◆ Renouf (1997) adds a fourth aim which is teaching to establish resources; (involves the learners in data collection, corpus design and corpus compilation)

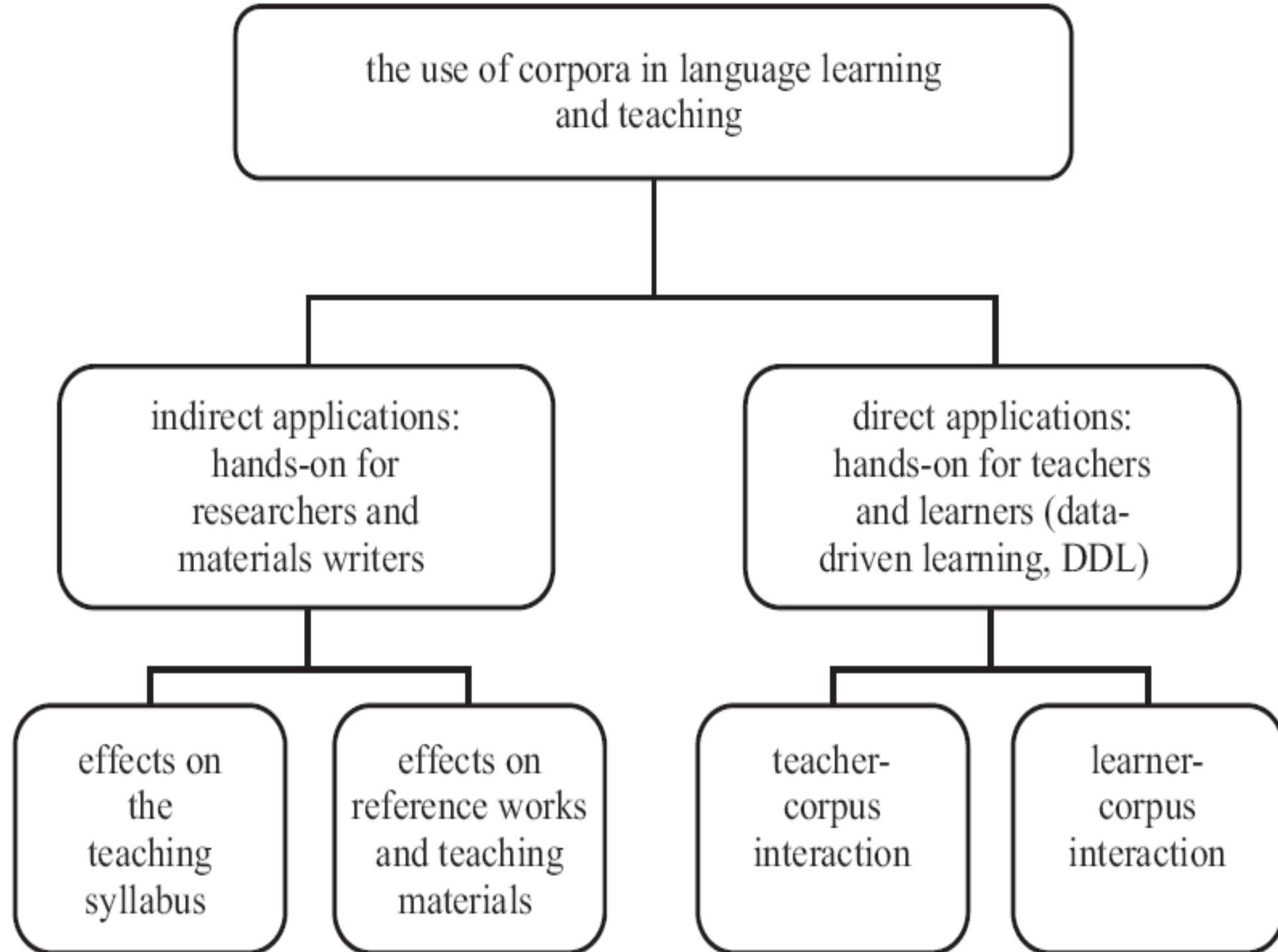


Fig. 7.2: Applications of corpora in language teaching



# Corpus-driven

- เปลี่ยนแปลงวิธีการเรียนการสอนไปเป็นแบบที่ Tim Johns เรียก data-driven learning (DDL)
- บางคนใช้ corpus-informed แทน corpus-driven
- ผู้สอนและผู้เรียนจะต้องเปลี่ยนแปลงบทบาทตัวเองให้เหมาะสมกับวิธีการใหม่
- ต้องมีการคัดเลือกตัวข้อมูลภาษาที่จะนำมาใช้
- ต้องเตรียมข้อมูลให้มีปริมาณที่มากพอสำหรับนำมาอนุมานกฎเกณฑ์ของการใช้ภาษาได้ เปลี่ยน deductive -> inductive
- ไม่มี pre-set explanation แล้วหาดย.มา fit แต่หาดย.จริงแล้วอธิบาย

# Corpus-driven

- ข้อดี : ผู้เรียนสามารถเข้าถึงข้อมูลได้โดยตรง
- มิใช่เรียนจากสิ่งที่สรุปในตำรา ซึ่งมักจะมีข้อเสียคือ ตำรามักจะเขียนขึ้นจากกฎเกณฑ์ต่างๆ เท่าที่จะสามารถนึกคิดเอาได้ และถ่ายทอดสืบกันไป เมื่อเวลาผ่านไปนานๆก็อาจจะไม่ได้คัดเลือกเอารูปภาพที่ไม่ใช้แล้วออกไป
- ทำให้ผู้เรียนจะไม่สามารถแยกแยะรูปภาพที่เป็นรูปแบบที่ใช้ปกติในปัจจุบัน (common use) กับรูปภาพที่เลิกใช้กันแล้ว (out of use) ได้

# Corpus-driven

- กฎที่ผู้เรียนค้นหาเองจะเข้าใจ จดจำได้ง่ายกว่าที่ได้จากการสอน
- ช่วยตอบคำถามที่ intuition ตอบได้ยาก เช่น  
disinterested, uninterested ต่างกันอย่างไร ลองค้น BNC

# Disinterested or Uninterested

- Disinterested means "impartial" or "not taking sides." (In other words, not having a personal interest at stake.)
- Uninterested means "not interested." (In other words, not showing any interest.)
  - Correct: A good referee should be disinterested.
  - (He does not take sides.)
  - Incorrect: He was disinterested in Jill's hobby.
  - Correct: He was uninterested in Jill's hobby.
  - (He shows no interest.)
  - <http://englishplus.com/grammar/00000207.htm>

# Disinterested or Uninterested

- If you are **disinterested** in something you are impartial and do not take sides: *A disinterested observer of the scene would have wondered what all the fuss was about.*
- If you are **uninterested** you have no interest at all: *The player was uninterested in the public reaction to his remark.*  
*Disinterested* is often used instead of *uninterested* to mean lacking interest. This use is widely regarded as incorrect and should be avoided, especially in formal writing.
- <http://www.tiscali.co.uk/reference/dictionaries/english/data/d0081934.html>

# Corpus-driven

- กฎที่ผู้เรียนค้นหาเองจะเข้าใจ จดจำได้ง่ายกว่าที่ได้จากการสอน
- ช่วยตอบคำถามที่ intuition ตอบได้ยาก เช่น  
disinterested, uninterested ต่างกันอย่างไร ลองค้น BNC
- ใช้ข้อมูลอะไร
  - ข้อมูลต้องเป็น authentic แค่นั้น?
  - ข้อมูลที่ผิดจากที่สอนในตำราไวยากรณ์จะใส่ใน corpus ไหม?
  - ถ้าคัดเลือกหรือ edit text ก็เป็น teacher control

# Corpus-driven

- การใช้คลังข้อมูลภาษาสำหรับการเรียนการสอนภาษานั้น จำเป็นต้องมี โปรแกรมคอนคอร์แดนซ์และคลังข้อมูลภาษาที่เหมาะสม
- ผู้เรียนจะมีบทบาทเสมือนเป็นนักวิจัยทางภาษา อาจารย์ผู้สอนเองก็จะเปลี่ยนบทบาทจากผู้ถ่ายทอดความรู้ไปเป็นผู้ประสานงานคอยให้ความช่วยเหลือ
- ซึ่งจะทำให้ก็ต้องอาศัยความพร้อมในหลายๆด้าน
- ผู้สอน & ผู้เรียน : ใช้ concordance ได้ มองหา generalization ได้ และมี corpus ที่เหมาะสมกับความต้องการ

# Corpus-driven

- Concordance ชวนเห็น pattern การใช้ชัดเจน เพราะเป็นการดึงเฉพาะข้อมูลที่น่าสนใจออกมาให้เห็นพร้อมกัน เห็นว่าอะไรเกิดซ้ำๆ
- นำสู่ lexical syllabus, pattern grammar, ready-made combination ไม่เน้นแยกชัดเจนระหว่าง lexical & grammar แบบเก่า
- มีทั้ง idiom principle และ open choice principle (Sinclair)
- ทำให้ประเด็นเรื่อง collocation, lexical phrase เป็นที่สนใจในการสอน



# Corpus-based

- คลังข้อมูลภาษาเป็นแหล่งข้อมูลที่สามารถอธิบายปรากฏการณ์ต่างๆ ในภาษาได้ดีขึ้น ช่วยให้สามารถเขียนพจนานุกรมหรือตำราไวยากรณ์ได้ดีขึ้นอันจะเป็นประโยชน์ต่อการเรียนการสอนภาษาอีกทอดหนึ่ง
- ช่วยให้สามารถวางหลักสูตรแนวการสอนได้ดีขึ้นเพราะสามารถเลือกคำศัพท์และโครงสร้างไวยากรณ์ที่เหมาะสมสำหรับการเรียนในลักษณะที่สอดคล้องกับการใช้ภาษาจริง
- ช่วยให้เห็นลักษณะที่แตกต่างกันของตัวบทแต่ละประเภท ใช้ประกอบการสอน ESP

# Corpus-based

- การนำคลังข้อมูลภาษามาใช้เลยไม่เหมาะ ผู้วางหลักสูตรต้องเลือกว่าผลที่ได้จากการศึกษาผ่านคลังข้อมูลแบบไหนที่สามารถนำมาใช้ให้เหมาะสมกับหลักสูตรที่วางไว้
- Aston (1995) เสนอให้จัด task ของการวิเคราะห์คลังข้อมูลให้เป็นงานหนึ่งใน communicative task ทำได้ 2 แบบ
  - ใช้คลังข้อมูลภาษาเพื่อเป็น reference tool สำหรับใช้แก้ปัญหาที่เกิดขึ้นในแบบฝึกหัด (communicative task) นั้นๆ เช่น ใช้คลังข้อมูลเฉพาะสาขาที่รวบรวมไว้เพื่อเป็นตัวอย่างเวลาที่สอนแปลเอกสารในลักษณะเดียวกัน
  - ใช้คลังข้อมูลภาษาเป็นแหล่งข้อมูลสำหรับการทำแบบฝึกหัด

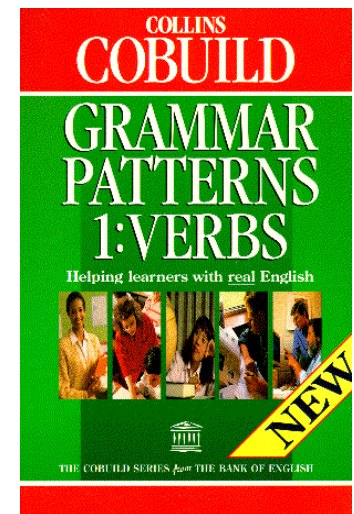
# Corpora in ELT















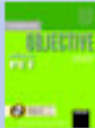
























- การใช้ Learner corpus (งานนักเรียน)
  - Concordance ช่วยให้ผู้เรียนเห็นภาพการใช้ผิดของตนเองที่ชัดเจนขึ้น
  - เมื่อนำเทียบกับการใช้ของ native ใน reference corpus จะทำให้นักเรียนเข้าใจที่ผิดและจดจำดีกว่า แก้ปัญหา fossilized pattern ใน interlanguage
- การใช้ Learner corpus (งานคัดเฉพาะเพื่อการสอน)
  - เลือกเฉพาะ text ที่ตรงประเด็นที่ต้องการสอน ให้ความหมายต่างๆ ชัดเจน ง่ายกว่าสำหรับนักเรียนในการหาคำตอบเอง
  - ช่วยในการเรียนรู้คำได้ดีและยั่งยืนกว่าการสอนแบบเดิม

# Corpora in ELT



- Corpus based : use corpus as a resource
  - References :
    - create better references
    - dictionary, grammar book, textbooks
  - verify certain hypotheses about languages
  - find support examples / counter examples
  - use a native corpus as a reference
  - see whether it is possible
  - which one is more natural



Level	Adult	Secondary	Exams	Professional English	Cambridge Copy Collection
Beginner					
Elementary	 	 	 		
Pre-Intermediate	 		  		
Intermediate	 		       	 	
Upper Intermediate					
Advanced			     		
Proficiency			 		

[http://www.cambridge.org/elt/corpus/corpus based books.htm](http://www.cambridge.org/elt/corpus/corpus%20based%20books.htm)

Level

Grammar

Vocabulary

Pronunciation

Dictionaries

Methodology and Linguistics

Beginner



Elementary



Pre-Intermediate



Intermediate



Upper Intermediate



Advanced



# Corpora in ELT

- Corpus based : use corpus as a resource
  - Syllabus design :
  - Can use both native and learner corpora
  - Native corpora => what are actually used
  - Learner corpora => what are the problems
  - Find out which aspects should be given priority

# Corpora in ELT

- Corpus based : use corpus as a resource
  - Syllabus design :
  - Lexical syllabus  
focus on frequency of occurrence
  - How many words the students should know?  
What are they?
  - Knowing 90% or 95% of the words?  
Can we ignore low frequency/difficult words?



Concordancer - [Word Frequency]					
File Edit Tools Window Help					
Word	F	F %	CF	CF %	
1. the	29485	3.528	29485	3.53	
2. to	26651	3.189	56136	6.72	
3. and	24798	2.967	80934	9.68	
4. of	23671	2.832	104605	12.52	
5. i	18090	2.164	122695	14.68	
6. a	14872	1.779	137567	16.46	
7. her	14346	1.716	151913	18.18	
8. in	12498	1.495	164411	19.67	
9. was	12063	1.443	176474	21.12	
10. it	11042	1.321	187516	22.44	
11. she	10858	1.299	198374	23.74	
12. not	9368	1.121	207742	24.86	
13. that	8940	1.070	216682	25.93	
14. be	8871	1.061	225553	26.99	
15. you	8624	1.032	234177	28.02	
16. he	7970	0.954	242147	28.97	
17. had	7898	0.945	250045	29.92	
18. as	7776	0.930	257821	30.85	
19. for	7401	0.886	265222	31.73	
20. with	6668	0.798	271890	32.53	
21. his	6497	0.777	278387	33.31	
22. but	6456	0.772	284843	34.08	
23. have	5861	0.701	290704	34.78	
24. is	5390	0.645	296094	35.43	
25. at	5312	0.636	301406	36.06	
26. s	4704	0.563	306110	36.63	
27. locdoc	4401	0.527	310511	37.15	
28. loc	4396	0.526	314907	37.68	
29. all	4228	0.506	319135	38.18	
30. so	4211	0.504	323346	38.69	
31. very	4178	0.500	327524	39.19	
32. on	4083	0.489	331607	39.68	
33. by	4003	0.479	335610	40.16	
34. my	3999	0.478	339609	40.63	
35. him	3982	0.476	343591	41.11	
36. could	3852	0.461	347443	41.57	

# Corpora in ELT

- Corpus based : use corpus as a resource
  - Materials design :
  - use authentic texts for reading exercises
  - use a corpus as sources of communicative tasks
  - concordance output =>
  - gapfill exercise
  - study of polyseme e.g. round
  - study of synonyms e.g. above vs over
  - jigsaw reading

# Gapfill Exercise

**Exercise 1** Use the context clues to identify the missing keyword in the following list.

\*t several times, and yet, when I ask you to ..... a trip with me, to act ballast-  
\*the wireless operator on the grounds would ..... any message that came to him.  
\*t that his craft was working well, let Ned ..... charge and devoted him to see in  
\*d with a smile, "I think I should prefer to ..... my first ride in your larger sh  
\*ds could form a notion of what was about to ..... place, Tom had sent his craft t  
\*have a plan partly thought out, but it will ..... some time to finish it.  
\*eedy, and leaving Tom and Captain Weston to ..... the first watch, the others wen  
\*who naturally acted as guide, "as it would ..... too long to put together and be

The missing keyword is .....

# Polysemy

**Exercise 2** According to the data given below, how many parts of speech can the word *round* be? Also identify the contexts in which *round* in different part of speech occurs.

1	have turned him off, he coolly turned round and said that my father was a ge
2	outh where the sun shone all the year round and he could ride half the day w
3	lue sky; smooth and dimpled slopes as round as cones, bare but for the green
4	in a brown hat and blue dress, with a round face and snub nose went and bo..
5	t the neighborhood and Jo was looking round for help when Laurie, who had he
6	, I rather like it, for she winds one round her finger as softly and prettil
7	electrified him by throwing her arms round his neck, and crying out, with a
8	ch untied the shoes which he had hung round his neck gave way, and both fell
9	ade with a very full skirt and little round jacket, but embroidered by her o
10	Then they would select a long, round stick --which must be also perfe
11	s had an ingenious way of fastening a round stone to the end of a long handl
12	I can't induce him to give me a good, round sum of account--fifty thousand,
13	pestle another stone was a smooth and round surface at the bottom was used.
14	quickly, he saw Micky Maguire running round the corner of the street which t
15	me spread the table, set the children round the fire and fed them like so ma
16	h her broom she soon swept a path all round the garden, for Beth to walk in
17	descended from his throne, and coming round to the side opened the door and
18	y met with a troop of women collected round the well, talking and laughing.
19	self tossing on the ocean in a voyage round the world, when the sound of voi
20	l be crying so hard that the tick fog round you will obscure the prospect.
21	to last till the silver wedding came round , and invented three different k
22	we pass," said Jo, as the plate went round , and the ice began to melt out
23	is year the summer floods brought the round , brown, fruitful cones to my v

# Jigsaw reading

**Exercise:** The contexts after the word "such" have come out in the wrong order. Can you put them back in order so that the first part of each context matches the second part? Write a number in the brackets at the end of each context to show which ending goes with each beginning.

1	anned the match from that hour; and when	such as would visit him on his own terms.	(.....)
2	especially by the Churchills, as making	such comfort in the very easy distance of ran	(.....)
3	Saw the letter, and he says he never saw	such a handsome letter in his life.	(.....)
4	Ry favourable idea of the young man; and	such untoward circumstances, can excite.	(.....)
5	And then there was	such success has blessed me in this instance,	(.....)
6	With	such an opinion, in confirmation of his own,	(.....)
7	made him unfit for any acquaintance but	such .	(.....)
8	Knightley, comprehended many	such a pleasing attention was an irresistible	(.....)
9	respect which a harmless old lady, under	such an amazing match, was proved to have muc	(.....)

# Synonyms

... llenges, of course, but it really will be kind of **above** the fray of – which they're dealing with a lot ...  
... he Sailor's Act – why didn't the President rise **above** it and say, well, we're not going to talk that w ...  
... the environment and human health and all of the **above**. Today the President will be celebrating the c ...  
... , as you know from the 1991 demarkation, extends **above** the 36th Parallel and we've enforced that robustl ...  
... tary force in a way that represses minorities. **Above** all else, he should not do what he is apparently ...  
... as easy to impose, for example, a no-drive zone **above** the 36th – MCCURRY: If you think it ...  
... you think it is "easy" to impose a no-drive zone **above** that parallel you should go visit with some folk ...  
... s knee, pointing to the area of the kneecap, just **above** the kneecap – four or five inches above. There, ...  
... p, just above the kneecap – four or five inches **above**. There, you've got your photo for the evening new ...  
... . Mariano was pleased with that. She wants him, **above** all else, to rest up. The first thing he did tod ...  
... ent; it reaffirms principles of cooperation, and **above** all, our commitment to a secure, undivided, and ...  
... e two sides stressed their determination to rise **above** the differences through practical steps. The key ...  
... And Tokyo. MCCURRY: And all of the **above**. VOICE: Are those the ones that are g ...  
... r reelection. VOICE: I mean, over and **above**, it's one thing to be concerned about trying to ...  
... URRY: There's no veto that Russia would have **over** new future members. VOICE: But they wo ...  
... , cold struggle, called the Cold War, to prevail **over** totalitarianism. And now, as we look ahead to the ...  
... n 1996 is down, probably the lowest it's been in **over** three decades, I think, which will be good news f ...  
... 0 years ago to the vision that we seek to pursue **over** the decades ahead. We will leave Monday evenin ...  
... coordination, and, where possible as it emerges **over** time, joint decision-making and joint action bet ...  
... nd President Yeltsin clarify this current dispute **over** the meaning of the Charter, with the Russians sa ...  
... ians saying that it gives them a veto effectively **over** NATO decisions, and the U.S. saying, well, they ...  
... VOICE: Well, how do you keep it from spilling **over**, though? TALBOTT: Well, I don't think ...  
... ink there's any danger or way that it will spill **over**. The NAC is perfectly capable, and has demonstrat ...  
.. Russian democracy. And I think what's been proven **over** the last year of very painstaking work – the la ...  
... be talking about it besides the FTC once you get **over** there? BERGER: Not me. VOICE: ...

# Concordance in a class room

- ดู [http://www.ict4lt.org/en/en\\_mod2-4.htm](http://www.ict4lt.org/en/en_mod2-4.htm) Using concordance programs in the Modern Foreign Languages classroom (link)
- กล่าวถึงการใช้ concordance ในการสอนภาษาต่างประเทศ และ activity ต่างๆ เพื่อสอนเรื่องที่ต้องการ เช่น
  - Guess the mystery word
  - Beware of false friends
  - Use of prepositions
  - Syntax of adverbs
  - Students' own writing Etc.

# Word Lists

- Teachers should teach materials which are relevant to the learners
- Teachers should teach the most useful vocabulary
- Teachers should teach the most frequent items first
- General Service List (West 1953) : 2000 คำใช้บ่อย  
คัดจาก corpus 5 ล้านคำ
- University Word List (Xue and Nation 1984) :  
ประมาณ 800 คำแยกเป็นระดับ 11 ระดับ ปัญหา corpus  
ที่ใช้มีขนาดเล็กเป็นทางวิทยาศาสตร์เทคโนโลยีมาก



# Word Lists

- Academic Word List (Coxhead 2000) เทียบกับ UWL ได้ corpus 3.5 ล้านคำ academic text 4 กลุ่ม 28 สาขา, เลือกคำที่พบอย่างน้อย 100 ครั้ง และพบอย่างน้อย 15 สาขา อย่างน้อย 10 ครั้งทั้ง 4 กลุ่ม
- ดู GSL UWL AWL
- ใช้ GSL + AWL เพื่องาน ELT ได้อย่างไร
- Download Range Program จาก Paul Nation's web <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- โปรแกรมตรวจสอบคำใน text ต่างๆ (มากถึง 32 files) พร้อมกับเทียบกับคำใน GSL/AWL หรือใน BNC wordlist

# Learner Corpus

- ◆ คลังข้อมูลภาษาผู้เรียน (learner corpus) เป็นคลังข้อมูลภาษาประเภทหนึ่งที่เก็บข้อมูลภาษาที่ได้จากผู้เรียนภาษาต่างประเทศ อาจเป็นภาษาเขียนหรือเป็นภาษาพูดก็ได้
- ◆ สะท้อนความไม่สมบูรณ์มากกว่าจะเป็นข้อมูลใช้อ้างอิง
- ◆ ทำให้เข้าใจสภาพการใช้ภาษาที่ผู้เรียนภาษานั้นเป็นตลอดจนปัญหาที่เป็นอยู่ว่ามีความเบี่ยงเบนจากภาษาของ native อย่างไร
- ◆ ทำให้เห็นภาษาในระหว่าง (interlanguage) ที่เกิดขึ้นในระหว่างพัฒนาการเรียนรู้อาษา

# Learner corpus

- ◆ ข้อมูลภาษาอะไรก็ได้ที่ผู้เรียนมีภาษาแม่เป็นอีกภาษาหนึ่ง
- ◆ เมื่อเต็ม เรียนภาษาอะไร ผู้เรียนเป็นผู้พูดภาษาอะไร พบว่ามีการใช้คำต่างๆ
- ◆ \_\_\_\_ corpus of learner English เช่น Greek corpus of learner English
- ◆ corpus of \_\_\_\_ learner English เช่น Corpus of Japanese Learner English
- ◆ corpus of \_\_\_\_ learner of English เช่น Corpus of Chinese Learner of English

- ◆ \_\_\_\_ learner English corpus เช่น Polish Learner English Corpus
- ◆ \_\_\_\_ English learner corpus เช่น Thai English learner corpus
- ◆ English learner corpus of \_\_\_\_ เช่น English learner corpus of Turkish students' Writings
- ◆ \_\_\_\_ learner corpus of English เช่น Taiwanese Learner Corpus of English
- ◆ \_\_\_\_ learner of English Corpus เช่น Santiago University Learner of English Corpus
- ◆ Corpus of English by \_\_\_\_ learners เช่น Corpus of English by Japanese Learners

# Learner corpus

- ◆ จะใช้ X learner English corpus เพราะใช้แพร่หลายกว่ารูปแบบอื่น
- ◆ learner corpus จะใช้ “คลังข้อมูลภาษาผู้เรียน” ไม่ใช่ “คลังข้อมูลผู้เรียนภาษา” เพราะเน้นความเป็นข้อมูลภาษา
- ◆ English learner corpus ใช้ “คลังข้อมูลผู้เรียนภาษาอังกฤษ” ไม่ใช่ “คลังข้อมูลภาษาของผู้เรียนภาษาอังกฤษ” เพื่อความกระชับ
- ◆ Thai English learner corpus ใช้ “คลังข้อมูลภาษาอังกฤษของผู้เรียนชาวไทย” คลังข้อมูลผู้เรียนภาษาอังกฤษที่เป็นชาวไทย

# Learner corpus

- ◆ เริ่มมีการประชุมวิชาการนานาชาติปี 1999 ที่ฮ่องกง โดย Sylviane Granger และ Joseph Hung
- ◆ งานวิจัยด้านคลังข้อมูลภาษาผู้เรียนมีการวิเคราะห์เชิงปริมาณแบบภาษาศาสตร์คลังข้อมูล มีการกำกับข้อมูลภาษากำกับข้อผิดพลาดชัดเจน มีการวิเคราะห์เปรียบเทียบต่าง (contrastive analysis) และวิธีการวิเคราะห์ข้อผิดพลาด (error analysis) แบบเดียวกับงานด้านการสอนภาษาต่างประเทศหรือภาษาที่สอง เพื่อหาปัญหาและความต่างจากภาษาของเจ้าของภาษา
- ◆ ที่ผ่านมามีการศึกษา error analysis แต่ใช้ข้อมูลที่เก็บมาแบบจำกัด ไม่สามารถให้คำตอบในภาพรวมได้

# Learner corpus

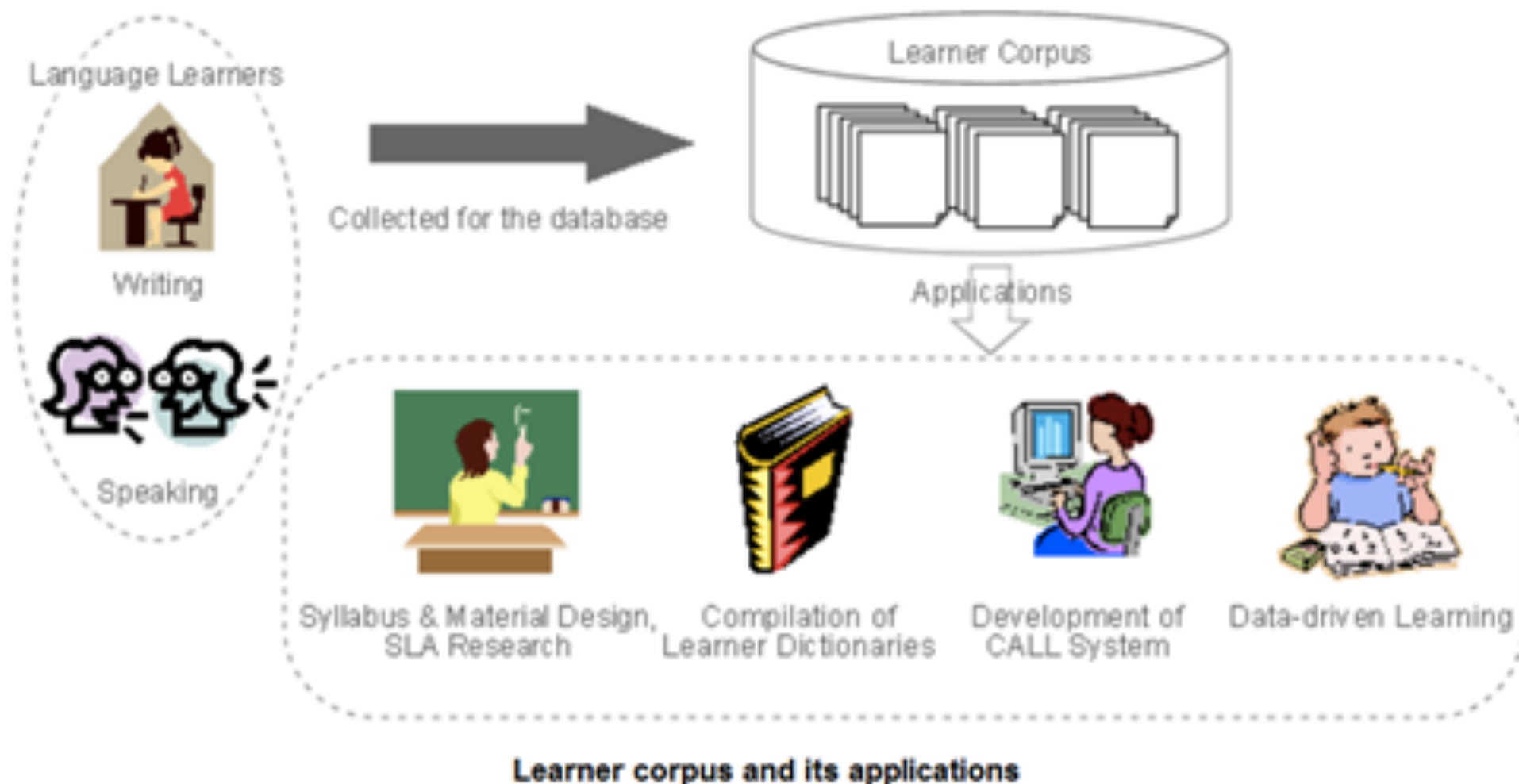
- ◆ เช่น พบว่านักเรียนมีการใช้คำบางคำผิด ก็อาจจะมีคำถามตามมาว่า
- ◆ คำที่ใช้ผิดนี้ ในที่อื่นๆ หรืองานเขียนลักษณะอื่น นักเรียนใช้ผิดอย่างสม่ำเสมอหรือมีที่ใช้ถูกผสมด้วย
- ◆ ข้อผิดที่พบเป็นเรื่องเฉพาะตัวหรือเป็นปัญหาร่วมของนักเรียนคนอื่นๆ ด้วย
- ◆ เมื่อมีความรู้ภาษาอังกฤษในระดับสูงขึ้น ข้อผิดนี้ยังคงพบอยู่หรือไม่
- ◆ ข้อผิดนี้พบในงานเขียนของเจ้าของภาษาด้วยหรือไม่ เจ้าของภาษามีการใช้คำนี้อย่างไร แตกต่างจากผู้เรียนอย่างไร
- ◆ จะตอบคำถามเหล่านี้ได้ต้องอาศัยข้อมูลมากกว่างานเขียนที่อาจารย์ผู้สอนเก็บจากนักเรียนของตนเองเท่านั้น
- ◆ ต้องอาศัยข้อมูลภาษาจากผู้เรียนจำนวนมาก

# ประเด็นการสร้าง Learner corpus

- ◆ ICLE (International Corpus of Learner English)  
(Granger 2003) กล่าวถึงประเด็นต่างๆ
- ◆ ขนาด : มากกว่าเก็บในชั้นเรียน ปัจจุบันเป็นหลักล้านคำ
- ◆ ออกแบบ : คำนี้ตัวแปรต่างๆ ที่เกี่ยวข้องและมีผลต่อภาษา
  - ◆ ตัวแปรผู้เรียน : อายุ เพศ ภาษาแม่ ...
  - ◆ ตัวแปร text : หัวข้อ ความยาว ...
  - ◆ ตัวแปร L2 : ประสบการณ์ภาษา ภาษาอื่นที่รู้ ...
  - ◆ ตัวแปรงาน : เวลาที่ให้ สอบ/การบ้าน เปิดหนังสือได้ ...



# ประโยชน์ learner corpus



# ประโยชน์ learner corpus

- ◆ Error analysis
  - ◆ มีการศึกษามานาน ช่วง 1970 ศึกษาข้อผิดพลาด
  - ◆ ข้อมูลมีจำนวนน้อย เก็บไม่เป็นระบบ ทำซ้ำยาก ไม่สะท้อน interlanguage น่าเชื่อถือพอ
  - ◆ การวิเคราะห์ subjective ตัดสินยากว่าผิดประเด็นไหน (แก้ถูกได้หลายแบบ)
  - ◆ สนใจแต่ข้อผิดพลาด ไม่เห็น over use, under use
- ◆ ข้อผิดพลาด collocation (Chí et al. 1994; Lorenz 1997; Granger 1998; Chen 1998) ข้อผิดพลาด connectors (Milton and Tsang 1993; Granger and Tyson 1996; Satoh and Fang 1998) ข้อผิดพลาด irregular past tense (Tono and Aoki 1998) ข้อผิดพลาด article (Mason and Uzar 2000)

# ประโยชน์ learner corpus

- ♦ contrastive analysis
  - ♦ เดิมดูจากลักษณะไวยากรณ์ว่ามีอะไรต่างกัน L1-L1 อาศัย intuition
  - ♦ ปัจจุบัน อาศัย corpus เปรียบเทียบ ไม่จำกัดที่ L1-L1 เทียบ L1-L2, L2-L2 ก็ได้
  - ♦ Granger เรียก Contrastive Interlanguage Analysis
  - ♦ เมื่อเทียบคลังข้อมูลผู้เรียนภาษากับคลังข้อมูลเจ้าของภาษา ก็จะ  
สามารถศึกษาพัฒนาการการเรียนรู้ภาษาที่หนึ่งเทียบกับภาษาที่  
สองในแง่มุมต่างๆ ว่าแตกต่างจากภาษาอังกฤษต้นแบบอย่างไร  
บ้าง โดยมีพื้นฐานความคิดว่าความแตกต่างเหล่านี้คือปัญหาของการ  
เรียนภาษาอังกฤษ

# ประโยชน์ learner corpus

- ◆ contrastive analysis
  - ◆ มีคำถามภาษาอังกฤษต้นแบบ=? US? British? ..
  - ◆ บางคนมองสิ่งที่พบใน learner corpus ไม่ใช่ error แต่เป็นผลจากวัฒนธรรมของแต่ละแห่ง
  - ◆ มีการศึกษาว่า ถ้า L1 ต่างกันมีผลต่อการเรียน L2 ภาษาอังกฤษอย่างไร
  - ◆ มี universal error? L1-specific error

# ประโยชน์ learner corpus

- ◆ performance analysis
  - ◆ ดู interlanguage
  - ◆ Leńko-Szymańska (2000) ศึกษาพัฒนาการจำนวนวงศัพท์ในผู้เรียน โดยศึกษาจากคลังข้อมูลภาษาผู้เรียน
- ◆ การจัดการเรียนการสอน
  - ◆ ทำให้ออกแบบการสอนที่เหมาะสม
  - ◆ สร้างทรัพยากรสำหรับผู้เรียน พจนานุกรม ตำรา

# ประโยชน์ learner corpus

**LONGMAN Dictionary of Contemporary English**

DICTIONARY ACTIVATOR EXERCISES GRAMMAR web/email POP UP

hardly OK search tools ?

**NEW**

WRITING ASSISTANT  
SPOKEN EXAMPLES  
TEACHER'S LESSONS

My boss **could hardly** be described as handsome.  
**hardly the time/place/person etc** (=a very unsuitable time, place, person)  
This is **hardly** the place to discuss the matter.

3 used to say that something has only just happened:  
The building work has **hardly** begun.  
**hardly ... when/before**  
She had **hardly** sat down when the phone rang.

**GRAMMAR**

Do not use **hardly** with a negative word

- I can **hardly** believe he said that (NOT I can't hardly believe he said that).
- There's **hardly** any milk left (NOT There's hardly no milk left).

Use **hardly** just before the main verb

- He **could hardly** speak (NOT He hardly could speak).

Do not use **hardly** at the beginning of a sentence, except in very formal writing

- I **had hardly** got in the house when the phone rang

is the usual way to say this. It is possible to say

- Hardly** had I got in the house when the phone rang

but this is very formal

!! Do not use **hardly** as the adverb of **hard**. The adverb of **hard** is **hard**

- I tried **hard** to remember (NOT I tried hardly to remember).
- Students have to study very **hard** (NOT Students have to study very hardly).

**WORD CHOICE**

**Phrase bank**

Dictionary phrases:

- can **hardly**
- could **hardly**
- hardly** a day/week/month etc goes by without/when

**Examples bank**

Extra dictionary examples

- All these similarities could **hardly** be due to chance.
- He's **hardly** a world chess champion - you should be able to beat him.

**Activate your language**

**a small amount of food or drink**

a little • a little bit • not much • very little • hardly/barely any • drop • pinch • dash

# การพัฒนา learner corpus

- ◆ มีการวางแผนออกแบบคลังข้อมูลที่ต้องการว่าควรมีลักษณะอย่างไร
- ◆ ต้องให้ข้อมูลที่ชัดเจนว่าเก็บข้อมูลใดมาบ้าง แจกแจงปัจจัยต่างๆ
- ◆ ปัจจัยเรื่องของภาษา
  - ◆ mode เป็นภาษาพูดหรือภาษาเขียน
  - ◆ genre ประเภทของตัวบท เช่น เป็นความเรียง จดหมาย บันทึกประจำวัน เรื่องสั้น เป็นต้น
  - ◆ style รูปแบบการเขียน เช่น เรื่องเล่า การอภิปราย
  - ◆ topic หัวข้อเรื่อง เช่น เรื่องทั่วไป เรื่องนันทนาการ

# การพัฒนา learner corpus

- ◆ ปัจจัยเรื่องชิ้นงาน (task-related)
  - ◆ การรวบรวมข้อมูลว่าเป็นการเก็บหลากหลาย (cross-sectional) หรือเก็บต่อเนื่อง (longitudinal)
  - ◆ วิธีการเก็บข้อมูลว่ามีการเตรียมการ (prepared) หรือเก็บตามสะดวก (spontaneous)
  - ◆ การใช้ข้อมูลอ้างอิง (มีการใช้พจนานุกรม ใช้ต้นฉบับ)
  - ◆ เวลาที่ใช้ (จำกัด ไม่จำกัด)
  - ◆ ที่ที่ทำงาน (ทำในชั้นเรียน ทำเป็นการทำงาน)



# การพัฒนา learner corpus

- ◆ ปัจจัยผู้เรียน

- ◆ internal-cognitive (age, cognitive-style)
- ◆ internal-affective (motivation, attitude)
- ◆ L1 background
- ◆ L2 background (ESL หรือ EFL, และระดับชั้นศึกษา)
- ◆ L2 proficiency (standard test score)

# Error tagging

- ◆ การกำกับข้อผิดพลาดนั้นก็มีความหลากหลายแล้วแต่ความต้องการและแนวทฤษฎีที่ใช้
- ◆ การพัฒนา โปรแกรมกำกับข้อผิดพลาด โดยอัตโนมัติยังเป็นไปได้อยาก
- ◆ การกำกับข้อผิดพลาดจึงเป็นไปในลักษณะทำด้วยคนมากกว่า แต่ก็ป็นงานที่หนักและใช้เวลามาก ควบคุมให้สม่ำเสมอยาก
- ◆ Computer-aided Error Analysis โปรแกรมเพื่อช่วยให้มีการกำกับข้อผิดพลาดได้ง่ายขึ้น เป็นระบบเพราะใช้แท็กต่างๆ ที่ออกแบบไว้แล้วว่าจะระบุข้อผิดพลาดอะไรได้บ้าง

# Error tagging

- ◆ ข้อผิดพลาดที่พบในงานผู้เรียนสามารถแก้ไขได้ในหลายวิธี เช่น
- ◆ The policeman was in this corner whistle but it was too late  
สามารถแก้ไขใหม่ได้สามแบบ
  - ◆ The policeman who was in this corner whistled but it was too late
  - ◆ The policeman was in this corner and whistled but it was too late
  - ◆ The policeman in this corner whistled but it was too late
- ◆ การกำกับข้อผิดพลาดว่าเป็นข้อผิดพลาดอะไรจึงเป็นงานที่เป็นอัตวิสัย

# Learner corpus project

- ♦ Longman Learners' Corpus 10 ล้านคำ
- ♦ Cambridge Learner Corpus 16 ล้านคำ
- ♦ ICLE (2002)
  - ♦ สร้างคลังข้อมูลผู้เรียนภาษาอังกฤษ เพื่อศึกษาเปรียบเทียบภาษาอังกฤษของผู้เรียนชาติต่างๆ
  - ♦ argumentative essays ความยาว 500-1000 คำ
  - ♦ ข้อมูลมาจาก advanced learners นักศึกษาสาขาวิชาภาษาอังกฤษชั้นปีที่ 3 หรือ 4
  - ♦ เริ่มต้นในปีค.ศ.1990 โดย Granger University of Louvain-la Neuve ประเทศ Belgium
  - ♦ ในปีค.ศ.2002 ได้ออกเป็นซีดีรอมบรรจุคลังข้อมูลผู้เรียนภาษาอังกฤษจากผู้เรียนชาติต่างๆ ได้แก่ Bulgarian, Czech, Dutch, Finnish, French, German, Italian, Polish, Russian, Spanish, และ Swedish

# Learner corpus project

- ◆ ICLE (2002)
  - ◆ มีข้อมูลภาษาอังกฤษแบบอังกฤษและแบบอเมริกาเพื่อใช้เป็นคลังข้อมูลเทียบภาษา (LOCNESS - the Louvain Corpus of Native English Essays - British pupils' A level essays: 60,209 คำ, British university students essays: 95,695 คำ, American university students' essays: 168,400 คำ, รวม 324,304 คำ)
  - ◆ ปัจจุบันมีข้อมูล 3 ล้านคำ จากผู้พูดภาษาต่างๆ 21 ภาษา (มี 21 subcorpora) ได้แก่ Arabic, Bulgarian, Brazilian, Chinese, Czech, Dutch, Finnish, French, German, Greek, Italian, Japanese, Lithuanian, Norwegian, Pakistani, Polish, Portuguese, Romanian, Russian, Slovene, Spanish, South African, Swedish, และ Turkish

# Learner corpus project

- ◆ ICLE (2002)
  - ◆ กำกับหมวดคำโดยโปรแกรม ICLE Tagger ที่พัฒนาโดยทีมงาน TOSCA (Tools for Syntactic Corpus Analysis) University of Nijmegen
  - ◆ ผู้เรียนที่ให้ข้อมูลกับ ICLE จะต้องกรอกข้อมูลผู้เรียน เพื่อเป็นข้อมูลอ้างอิง
  - ◆ เขียนอภิปรายในหัวข้อเรื่องที่กำหนดให้นักเรียนเลือก เช่น "Crime does not pay", "Feminism has done more harm to the cause of women than good", "Pollution : a silent conspiracy"
  - ◆ ไม่กำหนดเวลาเขียน ใช้ทรัพยากรอ้างอิงได้ แต่จะต้องเขียนขึ้นเอง
  - ◆ งานเขียนที่เป็นงานสอบหรือที่เป็นงานในชั้นเรียนก็สามารถเก็บได้ แต่ไม่ควรเกิน 25% ของข้อมูล
  - ◆ ข้อมูลแต่ละภาษามีขนาดประมาณ 200,000 คำ เก็บจากผู้เรียน 300-400 คน

# Learner corpus project

- ◆ โครงการ USP Learner Corpus (Tagnin 2003)
  - ◆ เก็บข้อมูลผู้เรียนภาษาต่างๆ ได้แก่ อังกฤษ เยอรมัน สเปน โดยผู้เรียนทั้งหมดเป็นชาวบราซิล
  - ◆ เป็น multilingual learner corpus ที่ต่างจาก ICLE ที่เป็นคลังข้อมูลผู้เรียนภาษาอังกฤษจากนักเรียนชาติต่างๆ
  - ◆ USP learner corpus ช่วยตอบคำถามอย่างเช่น อะไรที่เป็นปัญหาร่วมกันไม่ว่าจะเป็นการเรียนภาษาต่างประเทศอะไร หรืออะไรที่เป็นปัญหาเฉพาะของนักเรียนบราซิล
  - ◆ ICLE จะช่วยตอบคำถามอย่างเช่น อะไรที่เป็นปัญหาร่วมกันในการเรียนภาษาอังกฤษไม่ว่าจะเป็นนักเรียนชาติใด

# Learner corpus project

- ◆ โครงการ Multimedia Adult ESL Learner Corpus (MAELC) (Reder et al. 2003)
  - ◆ จัดทำโดยภาควิชา Applied Linguistics มหาวิทยาลัย Portland State University ร่วมกับ Portland Community College
  - ◆ จัดเก็บข้อมูลบันทึกเป็นวิดีโอของการเรียนการสอนภาษาอังกฤษเป็นภาษาต่างประเทศเป็นเวลาห้าปี (2001-2006)
  - ◆ มีข้อมูลประมาณ 5,000 ชั่วโมงมีนักเรียนผู้ใหญ่ (adult learner) จำนวนประมาณ 1,000 คน เป็นผู้เรียนระดับต่ำ (low level)



# Learner corpus project

- ◆ โครงการ Longman Learner's Corpus (LLC)
  - ◆ เพื่อประโยชน์เชิงพาณิชย์และในงานวิจัยสำหรับนักวิจัย
  - ◆ มีข้อมูลผู้เรียนภาษาอังกฤษชาติต่างๆ กว่า 10 ล้านคำ
  - ◆ ข้อมูลมีการกำกับหมวดคำ และเก็บข้อมูลผู้เรียนภาษาอังกฤษในระดับต่างๆ ซึ่งสามารถดูความสามารถทางภาษาอังกฤษของกลุ่มผู้เรียนต่างๆ
  - ◆ เป็นประโยชน์โดยตรงต่อการวิเคราะห์ข้อผิดพลาดและความสามารถที่มี ทำให้ผู้ที่ทำพจนานุกรม ผู้ที่ทำหนังสือเรียนต่างๆ กำหนดเนื้อหาและคำอธิบายการใช้ที่เหมาะสมกับผู้เรียนได้

# Learner corpus project

- ◆ โครงการ Cambridge Learner Corpus (CLC)
  - ◆ เก็บข้อมูลผู้เรียนชาติต่างๆ ทั่วโลก และมีไว้เพื่อใช้ภายในบริษัท
  - ◆ มีขนาดกว่า 30 ล้านคำจากนักเรียน 95,000 คนจาก 190 ประเทศและมีภาษาแม่ต่างๆ 130 ภาษา มีการกำกับข้อผิดพลาดด้วยระบบที่สำนักพิมพ์ Cambridge กำหนดขึ้น
  - ◆ CLC เป็นส่วนหนึ่งของคลังข้อมูล Cambridge International Corpus (CIC) สร้างโดยสำนักพิมพ์ Cambridge University Press และ ESOL
  - ◆ ข้อมูลภาษาผู้เรียนจึงมาผู้ที่สอบข้อสอบ Cambridge ESOL ทั่วโลก
  - ◆ ผู้ที่ทำพจนานุกรม หนังสือเรียนต่างๆ ให้กับสำนักพิมพ์ Cambridge สามารถเข้าถึงคลังข้อมูลนี้ได้ เพื่อใช้ศึกษาข้อผิดพลาดและความสามารถทางภาษาของผู้เรียนภาษาอังกฤษจากที่ต่างๆ ข้อมูล

# Learner corpus project

- ◆ โครงการ HKUST Corpus
  - ◆ ขนาดใหญ่แต่มีไว้ใช้ภายในองค์กร มีข้อมูลผู้เรียนกว่า 25 ล้านคำ
  - ◆ ไม่มีเว็บไซต์โครงการ ข้อมูลมีการกำกับหมวดคำ
  - ◆ (ข้อมูลจาก <http://leo.meikai.ac.jp/~tono/lcorpuslist.html>)
- ◆ โครงการ SILS corpus
  - ◆ เป็นคลังข้อมูลผู้เรียนภาษาอังกฤษของนักศึกษาใน School of International Liberal Studies ที่มหาวิทยาลัย Waseda
  - ◆ นักศึกษามีทั้งที่เป็นชาวญี่ปุ่นและชาวต่างชาติจากที่ต่างๆ บางคนก็เป็นผู้พูดภาษาอังกฤษเป็นภาษาแม่ เป็นข้อมูลที่เก็บเพื่อประโยชน์ในการวิจัยการเรียนการสอนภาษาอังกฤษเป็นภาษาต่างประเทศ (Muehleisen 2006)

# Learner corpus project

- ◆ โครงการ Polish-English Language Corpus Research and Applications (PELCRA)
  - ◆ เก็บคลังข้อมูลภาษาหลายคลังข้อมูล หนึ่งในนั้นคือ Polish Learner English Corpus ซึ่งเป็นคลังข้อมูลสำคัญในโครงการ PELCRA
  - ◆ เก็บข้อมูลผู้เรียนระดับต่างๆ และในหลายๆการใช้งาน มีข้อมูลประมาณ 500,00 คำและจะมีการกำกับหมวดคำโดยโปรแกรม CLAWS ของ Lancaster University

# Learner corpus project

- ◆ โครงการ NICT Japanese Learner Corpus (JLE) Corpus
  - ◆ เก็บข้อมูลผู้เรียนภาษาอังกฤษชาวญี่ปุ่น จัดทำโดย National Institute of Information and Communications Technology
  - ◆ เป็นภาษาพูดที่เก็บจากตัวอย่าง 1,281 คน มีขนาด 1.2 ล้านคำ ที่ได้จากการสอบ English oral proficiency interview test นับรวมเวลาทั้งหมด 300 ชั่วโมง
  - ◆ มีการกำกับข้อมูลพื้นฐาน (basic tag) ซึ่งจะบอกถึงโครงสร้างตัวบทว่าใครเป็นผู้พูด บอกผลัด บอกการพูดซ้อน เป็นต้น และมีการกำกับข้อผิดพลาด (error tag) ในระดับคำและระดับไวยากรณ์ (lexical and grammatical errors)

# Learner corpus project

- ◆ โครงการ CU-TU Thai Learner English Corpus
  - ◆ สร้างคลังข้อมูลภาษาอังกฤษที่ได้จากงานเขียนของผู้เรียนชาวไทย
  - ◆ กลุ่มแรกคือผู้เรียนภาษาอังกฤษระดับมหาวิทยาลัยชั้นพื้นฐานไม่จำกัดคณะ ได้ข้อมูลจากนักศึกษามหาวิทยาลัยธรรมศาสตร์จำนวนกว่า 540,000 คำ จากนิสิตจุฬาลงกรณ์มหาวิทยาลัยจำนวนกว่า 320,000 คำ
  - ◆ กลุ่มที่สองเป็นผู้เรียนภาษาอังกฤษระดับมหาวิทยาลัยชั้นสูงโดยเลือกนิสิตเอกภาษาอังกฤษคณะอักษรศาสตร์ ได้ข้อมูลจำนวน 37,900 คำ
  - ◆ กลุ่มที่สามเก็บข้อมูลภาษาอังกฤษชาวไทยที่มีความรู้ภาษาอังกฤษดีเขียนภาษาอังกฤษประกอบวิชาชีพ ข้อมูลได้ 730,000 คำ

<http://www.arts.chula.ac.th/~ling/TLE/>

http://ling.arts.chula.ac.th/TLE/

Latest Headlines Nation Krungtheptura... WebMail - CU Gmail Facebook Linguistics User Identification Classic FM Bookmarks

## Thai Learner English Corpus

Faculty of Arts, Chulalongkorn University.

Search for <input type="text"/> Exact	Left context 60 Right context 60	Corpus : Intermediate Learner
Intermediate TLE	University All	Faculty All
Advanced TLE	University : CU	Faculty : Arts
Professional TLE	Newspaper All	Genre All
Sorted by : right context	Sample 5 samples	Concord

1. Thai English Learner corpus is the corpus of essays collected from undergraduate students from Chulalongkorn university and Thammasart university.
2. The corpus is categorized into two levels : intermediate learners and advanced learners.
3. Learners at the Intermediate level are first year students in different faculties.
4. Learners at the advanced level are second year students in the faculty of Arts, whose major is English.
5. Professional TLE are writings of Thai journalists in two English newspapers, the Nation and Bangkok Post.
6. Essays from Thammasart University's students were collected by Assist.Prof.Vilaivan Aroonmanakun, Language Institute.
7. Essays from Chulalongkorn University's students were collected by Assist.Prof.Prima Mallikamas, Ajarn Somjit Jirananthiporn, and other faculties in the English department.

# Research Topics

- Creating Learner Corpus :  
Design : subject, aim, variables, methodology, evaluation, etc.  
Is there a standard on error tagging?
- Comparing Corpus-based/Corpus-driven method and traditional method :  
How do we evaluate? Using pre-test, post-test?  
Evaluate what?
- Linguistic analysis on a learner corpus :  
error analysis, performance analysis, effect of L1 transfer, ...



# Research Topics

- Comparing different usages between a learner corpus and a native corpus (contrastive analysis)
- Comparing standard/general English and specific English
- Language variation : English as a global language  
eg. British vs American English
- Research on Language acquisition, Interlanguage

**Table 4.2: Summary of qualitative evaluation of concordancing**

Study	Subjects/Proficiency level	Aims of evaluation	Concordance activity	Data collection	Results
Hadley (forthcoming)	25 Japanese, first-year university students / Beginner	To elicit the learners' attitudes towards concordance-based activities	Concordance-based exercises in commercial textbook	Classroom observations	Once the students were familiar with concordances, they actively engaged in concordance-based tasks.
				Questionnaires	The majority of students said concordance-based exercises help improve their English.
Gan et al. (1996)	48 Malaysian undergraduate students / Possibly intermediate	To compare the learners' attitude towards concordance-based tasks and traditional methods as a means of teaching word in context	Guided concordancing (Self-concordancing following task guidelines)	Questionnaires	The majority of students preferred concordance-based tasks.
Blappert (1998)	Korean students / Possibly intermediate	To observe a DDL lesson following the process of identifying, classifying, and generalising concordance data	Concordance-based matching exercises and classroom discussion	Classroom observations	The class successfully followed the three steps. The students made useful generalisations and even questioned the teacher's categorisation of concordances and word uses.
Bernardini (1999)	Six Italian undergraduate students in translation / Possibly advanced	To survey the learners' experiences of independent corpus investigation	Independent concordancing (Self-concordancing of any language focus)	Classroom observation and learners' reports	The learners successfully adopted cognitive and linguistic strategies to interpret and generalise their findings, particularly in observing collocation and semantic prosodies of a search word.

Study	Subjects/Proficiency level	Aims of evaluation	Concordance activity	Data collection	Results
Turnbull and Burston (1998)	Two non-native English speaker postgraduate students at Monash University, Australia / Possibly advanced	To carry out a longitudinal case study of concordancing strategies used for the investigation of self-selected concordances.	Guided concordancing (Self- concordancing of words used wrongly by the students)	Interviews and unstructured conference sessions	The learners experienced varying levels of success with concordancing strategies, depending on cognitive style and motivation.
Maneeekhao (2001)	Thai postgraduate students / Possibly advanced	To assess the learners' performance in using self-selected concordances to correct errors	Guided concordancing (Self- concordancing of words used wrongly by the students)	Classroom observations	The majority of the students were able to induce simple rules and correct their mistakes, although most of them showed the search results to the teacher to ask for approval and often double-checked with peers before making correction.
Kennedy and Miceli (2001)	Eight undergraduate students in the Italian programme at Griffith University, Australia. / Possibly advanced	To evaluate the effectiveness of learners' independent investigations of Contemporary Written Italian Corpus (CWIC)	Independent concordancing (self-selected language focus based on general texts and learners' writing assignments)	Video-recording of learners' discussions and interviews	Although the students were capable of making some successful investigations, they still had problems with specifying searches, selecting relevant citations, and making valid generalisations.

**Table 4.3: Summary of quantitative evaluation of concordancing**

Study	Subjects/Proficiency level	Aims of evaluation	Research methods	Measurement	Results
Stevens (1991)	First -year university students at Sultan Qaboos University in Oman / Possibly intermediate	To test the hypothesis that concordance-based vocabulary exercises can be more easily solved than traditional gap-filler exercises, when testing learned vocabulary	Versioning (Two groups of students worked on both concordance and non-concordance exercises. Treatment was repeated at a later class)	Pair-matched <i>t</i> -test was used to compare mean scores between the two types of exercises.	The subjects received higher scores in concordance-based vocabulary exercises.
Gan et al. (1996)	48 Malaysian undergraduate students/Possibly intermediate	To compare the effectiveness of computer-based concordancing exercises and the conventional instructional approach in the teaching of words in context (e.g. prepositions, collocation)	Experimental design (Experimental group did guided concordancing, while control group was taught explicitly)	Pre- and post-tests (different test items) at the beginning and the end of the experiment	The subjects who learned through the computer concordancing approach performed better than those who went through conventional classroom instruction.
Cobb (1997)	11 university students at Sultan Qaboos University in Oman / Possibly intermediate	To test the hypothesis that concordancing has an advantage over traditional methods in helping students learn new vocabulary.	Versioning (Subjects working on the concordance and non-concordance CALL programs on alternate weeks for 12 weeks)	Identical pre- and post- tests on general vocabulary.	The score of the 11 subjects was increased by 21.5% in the post-test.
				Weekly quizzes (Spelling tests and novel-text tasks)	The mean score on weeks with concordancing was significantly higher on novel-text tasks.
Cobb (1999a) Cobb and Horst (2001)	58 university students at Sultan Qaboos University in Oman / Lower intermediate and upper intermediate	To test the effect of a concordance-based CALL on the learning of new vocabulary	Experimental design at two proficiency level (Experimental groups working on the concordance version, while the control groups working on the non-concordance version)	Identical pre- and post-tests (Definition tests and novel-text tasks)	At both levels, the experimental group performed significantly better in the novel-text tasks.



Study	Subjects/Proficiency level	Aims of evaluation	Research methods	Measurement	Results
Blappert (1998)	8 Korean students/Possibly intermediate	To test the learning effect of concordance-based materials focusing on the differences between verbs with similar meanings	Pre-experiment (No control group. All students did the same concordance-based tasks and discussions)	Pre- and post-tests (different test items) at the beginning and the end of the lesson	The average score of the students for the pre-test was 60%, while the post-test average was 87.5%.
Someya (2000)	40 Japanese businessmen / Possibly advanced	To test the effectiveness of a web concordancer in helping the learners write business letters with fewer errors	Experimental design (The experimental group was instructed to consult the web concordancer, while the control group did not have access to the concordancer)	Error counting (Preposition and article errors in the writing tasks submitted by both groups were counted)	The experimental group made significantly fewer article and preposition errors in the six tasks analysed.
Watson Todd (2001a)	25 postgraduate students at King Mongkut's University of Technology Thonburi, Thailand. / Possibly advanced	To investigate the ability of learners to induce rules and patterns from self-selected concordances and to use such induction to make self-correction of their errors.	Non-experimental (All subjects followed the same instructions but investigated different words)	First, concordance lines that correctly described or illustrated the learner's induced rules or patterns were counted. Then, successful corrections based on the induced rules were counted.	Generally, the students were able to induce valid patterns from their self-collected concordances and make valid self-corrections of their errors, and there was a strong correlation between these two abilities.

# ผลงานวิจัยช่วยวาง syllabus ที่เหมาะสม

Source	Language area	Findings
Carter and McCarthy 1995a, 1995b	<ul style="list-style-type: none"> <li>▪ Ellipsis</li> <li>• Tails used for reinforcement (e.g. <i>You're stupid, <u>you are</u>.</i>)</li> <li>• Left dislocation and topical information (e.g. <i><u>This friend of mine, her son</u> was in hospital ...</i>)</li> <li>• Indirect reporting with <i>say</i> and <i>tell</i> in past continuous</li> </ul>	A sampling of the items from a longer list of recurring features of spoken English found in the CANCODE corpus, which offers "great potential for new and revised description and pedagogy" (145).
Conrad 2000; Conrad and Biber 2000	<ul style="list-style-type: none"> <li>▪ Register-specific descriptions</li> <li>▪ Adverbial marking of stance</li> </ul>	"Corpus-based research has consistently shown that grammatical patterns differ systematically across" registers, for example, in the way adverbials are used as stance markers in conversation, journalism and academic writing (549).
Fox 1998	Vague language	Expressions such as <i>something, something like that, things like that</i> , etc. which are common features of unplanned discourse and yet noticeably absent from most classroom language (29).

Sinclair and Renouf 1988	“Delexicalized” verbs	(e.g. <i>make, do, put, take</i> , etc.) “A major feature of language which is not currently taught in textbooks” (153).
Willis 1990	<ul style="list-style-type: none"> <li>• The complex noun phrase</li> <li>• Discourse-structuring language</li> </ul>	These areas are found to be common in corpora but underrepresented in the conventional grammar syllabus, and thus deserve more attention.
Willis 1994	<ul style="list-style-type: none"> <li>▪ Second conditional</li> <li>• The passive</li> <li>• Reported speech</li> </ul>	Corpus research suggests that the attention given to these conventional syllabus components may be unwarranted. Willis suggests a more lexical approach to bring them in line with the data: for example, teaching <i>would</i> with hypothetical meaning by itself, teaching the passive adjectivally, or presenting specific verbs used to report speech or thought.

# Research Topics

- Linguistic analysis on a corpus : (corpus linguistics : Refinements and Reassessments 2009)
  - The interpersonal function of going to in written American English / Anna Belladelli
  - Re-analysing the semi-modal ought to: an investigation of its use in the LOB, FLOB, Brown and Frown corpora/ Marta Degani
  - On the use of split infinitives in English / Javier Calle-Martín and Antonio Miranda-García
  - Exploring change in the system of English predicate complementation, with evidence from corpora of recent English / Juhani Rudanko



# Research Topics

- Linguistic analysis on a corpus : (corpus linguistics : Refinements and Reassessments 2009)
  - Encoding of goal-directed motion vs resultative aspect in the COME + infinitive construction / Sara Gesuato
  - A corpus-based analysis of invariant tags in five varieties of English / Georgie Columbus
  - Discourse presentation in EFL textbooks: a BNC-based study / Christoph Rühlemann
  - Awful adjectives: a type of semantic change in present-day corpora / Göran Kjellmer

# Research based on learner corpus

- ◆ Flowerdew (1998) เปรียบเทียบคลังข้อมูลผู้เรียน HKUST กับ expert corpus MCC เพื่อศึกษาสิ่งที่เป็น cause กับ effect marker ว่ามีการใช้ผิด (misuse) ใช้มากเกินไป (overuse) ใช้น้อยเกินไป (underuse) อย่างไร
- ◆ Altenberg and Granger (2001) ศึกษาเปรียบเทียบการใช้คำที่มีความถี่การใช้สูง (high frequency word) ของผู้เรียนและเจ้าของภาษา โดยเลือกคำว่า make เป็นกรณีศึกษา
- ◆ ดูว่าผู้เรียนมีแนวโน้มที่ใช้คำกลุ่มนี้มากเกินไปหรือน้อยเกินไป มาจากการถ่ายโอนภาษาหรือไม่
- ◆ ข้อมูลภาษาผู้เรียนมาจาก ICLE เทียบกับคลังข้อมูลเจ้าของภาษา Louvain Corpus of Native English Essays (LOCNESS) พบว่าผู้เรียนมีการใช้คำ make ในลักษณะ delexical uses น้อยกว่าเจ้าของภาษา และใช้เป็น causative uses แตกต่างกันในกลุ่มผู้เรียน โดยนักเรียนชาวสวีเดนใช้เป็น causative uses มาก ในขณะที่นักเรียนชาวฝรั่งเศสใช้ในหน้าที่นี้น้อยกว่าเจ้าของภาษา

# Research based on learner corpus

- ◆ Sugíura (2002) ศึกษาความรู้เรื่องคำปรากฏรวมของผู้เรียนภาษาเปรียบเทียบกับเจ้าของภาษา
- ◆ Sugíura ตั้งคำถามเรื่องความรู้การใช้คำปรากฏรวม (collocational knowledge) ของผู้เรียนภาษาอังกฤษชาวญี่ปุ่น โดยศึกษาเปรียบเทียบกับส่วนที่ fixed expression ใช้วิธีให้เจ้าของภาษาเขียนประโยคใหม่ (paraphrase) จากข้อมูลที่ได้จากผู้เรียนทีละประโยค จากนั้นใช้โปรแกรม n-gram เพื่อดึงคำที่ติดกันออกมาเป็นข้อมูลของวลิตายตัว วิธีนี้ทำให้ได้ข้อมูลคำปรากฏรวมที่เทียบเคียงกันได้ชัดเจน ผลก็พบว่าผู้เรียนมีการใช้รูปแบบคำเฉพาะของตนแตกต่างจากเจ้าของภาษา

# Research based on learner corpus

- ◆ Neff et al. (2003) ศึกษาการใช้ modal verbs (can, could, may, might and must) และกริยา reporting verbs (suggest, wonder, argue, explain, express, recognise, say, show, and state) ในผู้เรียนภาษาอังกฤษจากพื้นภาษาแม่ต่างๆ ห้าภาษาได้แก่ Spanish, Dutch, Italian, French และ German โดยใช้คลังข้อมูล ICLE
- ◆ พบว่าผู้เรียนมีแนวโน้มใช้ modal verb ทั้งที่มากเกินไป (overuse) หรือใช้น้อยเกินไป (underuse) เมื่อเทียบกับงานเขียนโดยคนอเมริกันที่เป็นเจ้าของภาษา ในส่วนของกริยา reporting verb เจ้าของภาษาจะมีการใช้คำที่หลากหลายมากกว่าและมีหน้าที่ความหมายซับซ้อนกว่า

# Error tagging

- ◆ Dulay, Burt, and Krashen (1982) พูดถึงการจำแนกข้อผิดพลาดเป็นสองประเภท
  - ◆ ประเภทแรกเป็นข้อมูลภาษา อย่างเช่น การเติมวิภัติปัจจัยผิด การใช้คำผิด ข้อผิดพลาดทางไวยากรณ์ ข้อผิดพลาดการใช้กริยาช่วย การใช้คำบุพบท เป็นต้น
  - ◆ ประเภทที่สองเป็นข้อผิดพลาดทางรูป เช่น ตกคำ สะกดผิด สลับที่ เป็นต้น
- ◆ ปัจจุบัน การกำกับข้อผิดพลาดเป็นการกำกับแท็กลงบนคลังข้อมูลภาษาผู้เรียน ซึ่งจะมองรวมเป็นระบบเดียว การกำกับข้อผิดพลาดไม่ใช้การแก้ไขงานเขียนของนักเรียนให้สะอาดสวยงามหรือสมบูรณ์แบบ แต่เป็นการบันทึกว่ามีอะไรที่ผิดไปบ้าง

# Error tagging

- ◆ Granger (2003) กล่าวถึงระบบการกำกับข้อผิดพลาดที่ดีนั้น ควรมีคุณลักษณะสี่ประการ คือ
  - ◆ 1) ทำให้การกำกับข้อมูลเป็นไปแบบ consistent ไม่มีความแตกต่างกันเมื่อเปลี่ยนผู้ที่กำกับข้อผิดพลาด จะต้องมีการจัดทำคู่มือการกำกับข้อผิดพลาดที่ละเอียดและชัดเจน
  - ◆ 2) ข้อผิดพลาดที่กำกับนั้นควรให้ข้อมูลที่เพียงพอแก่ผู้ใช้สำหรับวิเคราะห์ข้อผิดพลาดเรื่องต่างๆ ได้ แต่ในขณะเดียวกันก็ต้องไม่มีข้อมูลที่มากเกินไปที่ทำให้จัดการได้ยาก (informative and manageable)
  - ◆ 3) มีความยืดหยุ่นเพียงพอที่จะเพิ่มเติมหรือลบแท็กต่างๆ ในระหว่างการกำกับข้อผิดพลาด และสามารถเรียกใช้งานได้ง่ายหลังจากที่กำกับข้อผิดพลาดเสร็จแล้ว (flexible)
  - ◆ 4) ชุดของแท็กต่างๆ ที่ใช้ควรจะใช้ได้กับข้อมูลชุดอื่นๆ หรือภาษาอื่นๆ ด้วย (reusable)

# Error tagging

- ◆ ระบบที่ใช้กำกับข้อผิดพลาดยังไม่ได้มีข้อตกลงที่เป็นมาตรฐานเดียวกัน
- ◆ Nicolls (2003) อธิบายถึงระบบการกำกับข้อผิดพลาดใน Cambridge Learner Corpus ว่าใช้รูปแบบการแท็กข้อผิดพลาดแบบ `<#CODE>wrong word|corrected word</#CODE>` ข้อผิดพลาดส่วนใหญ่จะใช้ตัวอักษรสองตัวในเครื่องหมายแท็กเพื่อบอกประเภทข้อผิดพลาด ตัวอักษรแรกบอกประเภทใหญ่ ตัวอักษรที่สองบอกประเภทย่อย

# CLC error tagging



รหัสตัวที่หนึ่ง

- F wrong Form used
- M something Missing
- R word or phrase needs Replacing
- U word or phrase is Unnecessary  
(i.e. redundant)
- D word is wrongly Derived

รหัสตัวที่สองบอก word class

- A Pronoun (Anaphoric)
- C Conjunction (linking word)
- D Determiner
- J Adjective
- N Noun
- Q Quantifier
- T Preposition
- V Verb (includes modals)
- Y Adverb (-ly)



# CLC error tagging

- ◆ He died <#MP>we|.We</#MP> buried him the next day.  
= ตก full stop ต้องแก้ไขด้วยการแทนที่ we ด้วย . We
  - ◆ MP = punctuation Missing
  - ◆ RP = punctuation needs Replacing
  - ◆ UP = Unnecessary punctuation
- ◆ ข้อผิดพลาดเกี่ยวกับ false friend กำกับด้วยรหัส FF แล้วตามด้วยรหัสบอก word class
- ◆ agreement error กำกับด้วยรหัส AG แล้วตามด้วยรหัสบอก word class
- ◆ มีรหัสสำหรับ error แบบต่างๆ อีกมาก

# CLC error tagging

- ◆ รหัส AS (incorrect Argument Structure) ใช้กรณีที่จำนวน argument ไม่เป็นไปตามที่ควร และไม่สามารถมองเป็นข้อผิดพลาดเรื่องตกคำอย่าง \*he explained me หรือมีคำเกินอย่าง \*he told to me
- ◆ รหัส AS ใช้กรณีเช่น ใช้ double object ผิด. \*it caused trouble to me จะถูกกำกับเป็น <#AS>it caused trouble to me|it caused me trouble</#AS>
- ◆ รหัส SA (American Spelling) ความจริงไม่ใช่ข้อผิดพลาด แต่กำกับไว้เพื่อประโยชน์ให้แน่ใจว่านักเรียนจะรู้ว่าตนเองใช้ภาษาอังกฤษตามแบบไหน
- ◆ รหัส SX (spelling confusion error) ใช้กำกับข้อมูลคำผิดที่ผู้เรียนสะกดผิดเป็นอีกคำที่คล้ายกัน เช่น to/too, their/there and weather/whether

# CLC error tagging

- ◆ การกำกับข้อผิดพลาดสามารถวางซ้อนกันได้
- ◆ I like to <#S>hospítelízel hospítalíse</#S> my relatives  
บอกว่าเป็นการสะกดคำผิด
- ◆ หากนักเรียนต้องการหมายถึง ‘put them up’ หรือ ‘have them to stay’ (show them hospítality) ดังนั้นจึงควรกำกับด้วยแท็ก RV ด้วย <#RV><#S>hospítelízelhospítalíse</#S>|put up</#RV>. หมายความว่า hospítelíze เป็นการสะกดผิด คำที่ถูกต้องคือ hospítalíze แต่จริงๆ ในประโยคนี้ควรใช้กริยา put up มากกว่า

# CLC error tagging

- ◆ ในการกำกับข้อผิดพลาด ผู้วิเคราะห์มักจะพบว่าวิธีแก้ที่ผิดได้มากกว่าหนึ่งวิธี เช่น \*He said me that ... สามารถแก้เป็น He said to me that หรือ He told me that
- ◆ คำถามคือจะเลือกกำกับข้อผิดพลาดอย่างไร หลักการคือพยายามรักษาแบบที่นักเรียนเขียนมากที่สุด
- ◆ ในกรณีนี้ การแก้ argument structure ของคำกริยาจะดูใกล้เคียงของเดิมมากกว่าการแก้คำกริยา จึงกำกับข้อผิดพลาดเป็น 'He said <#MT>|to</#MT> me that' มากกว่าที่จะเป็น 'He <#RV>said|told</#RV> me that' อีกทั้งการกำกับแบบแรกทำให้ผู้เรียนเห็นว่าจะใช้กริยา say ให้ถูกอย่างไร ดีกว่าสอนให้เลี่ยงการใช้กริยา say ไปใช้ tell แทน

# ICLE error tagging

- ◆ ระบบที่ใช้ใน ICLE ประกอบด้วยการกำกับข้อมูล 3 ระดับ error domain error category และ word category หรือหมวดคำ
- ◆ ไม่ได้แยกข้อผิดพลาดเป็นข้อผิดพลาดจริง (error) หรือเป็นเพียงข้อผิดพลาด (mistake) ให้เป็นเรื่องการวิเคราะห์ของแต่ละคนเอง
- ◆ โดเมนข้อผิดพลาดบอกหมวดใหญ่ซึ่งแยกย่อยเป็นประเภทข้อผิดพลาด ส่วนประเภทของคำกำกับไว้เพื่อสะดวกต่อการวิเคราะห์ เช่น ต้องการหาข้อผิดพลาดทั้งหมดที่เกี่ยวข้องกับคำบุพบท เป็นต้น
- ◆ ประเภทของคำจะกำกับเองด้วยมือเฉพาะส่วนที่ถูกกำกับว่าเป็นข้อผิดพลาด ไม่ได้ใช้โปรแกรม POS tagging เพราะไม่แน่ใจความถูกต้องเมื่อใช้กับงานเขียนผู้เรียนภาษา

# ICLE error tagging

Error Domains		Error Categories	
<F>	Form	<AGL> <MAJ> <DIA> <HOM> <GRA>	Agglutination Upper/lower case Diacritics Homonymy Other spelling errors
<M>	Morphology	<MDP> <MDS> <MFL> <MFC> <MCO>	Derivation-prefixation Derivation-suffixation Inflection Inflection-confusion Compounding
<G>	Grammar	<CLA> <AUX> <GEN> <MOD> <NBR> <PER> <TPS> <VOI> <EUF>	Class Auxiliary Gender Mode Number Person Tense Voice Euphony

# Error Tags for the NICT JLE Corpus (Izume et al. 2005)

Tag	Error category
<b>NOUN</b>	
<code>&lt;n_inf&gt;...&lt;/n_inf&gt;</code>	Noun inflection
<code>&lt;n_num&gt;...&lt;/n_num&gt;</code>	Noun number
<code>&lt;n_cs&gt;...&lt;/n_cs&gt;</code>	Noun case
<code>&lt;n_cnt&gt;...&lt;/n_cnt&gt;</code>	Countability of noun
<code>&lt;n_cmp&gt;...&lt;/n_cmp&gt;</code>	Complement of noun
<code>&lt;n_lxc&gt;...&lt;/n_lxc&gt;</code>	Lexis
<b>VERB</b>	
<code>&lt;v_inf&gt;...&lt;/v_inf&gt;</code>	Verb inflection
<code>&lt;v_agr&gt;...&lt;/v_agr&gt;</code>	Subject-verb disagreement
<code>&lt;v_fml&gt;...&lt;/v_fml&gt;</code>	Verb form
<code>&lt;v_tns&gt;...&lt;/v_tns&gt;</code>	Verb tense
<code>&lt;v_asp&gt;...&lt;/v_asp&gt;</code>	Verb aspect
<code>&lt;v_vo&gt;...&lt;/v_vo&gt;</code>	Verb voice
<code>&lt;v_fin&gt;...&lt;/v_fin&gt;</code>	Usage of finite/infinite verb
<code>&lt;v_ng&gt;...&lt;/v_ng&gt;</code>	Verb negation
<code>&lt;v_qst&gt;...&lt;/v_qst&gt;</code>	Question
<code>&lt;v_cmp&gt;...&lt;/v_cmp&gt;</code>	Complement of verb
<code>&lt;v_lxc&gt;...&lt;/v_lxc&gt;</code>	Lexis

---

## MODAL VERB

---

<mo\_lxc>...</mo\_lxc>

Lexis

---

## ADJECTIVE

---

<aj\_inf>...</aj\_inf>

Adjective inflection

<aj\_us>...</aj\_us>

Usage of positive/comparative/superlative of adjective

<aj\_num>...</aj\_num>

Adjective number

<aj\_agr>...</aj\_agr>

Number disagreement of adjective

<aj\_qnt>...</aj\_qnt>

Quantitative adjective

<aj\_cmp>...</aj\_cmp>

Complement of adjective

<aj\_lxc>...</aj\_lxc>

Lexis

---

## ADVERB

---

<av\_inf>...</av\_inf>

Adverb inflection

<av\_us>...</av\_us>

Usage of positive/comparative/superlative of adverb

<av\_pst>... </av\_pst>

Adverb position

<av\_lxc>...</av\_lxc>

Lexis

---

## PREPOSITION

---

<prp\_cmp>...</prp\_cmp>

Complement of preposition

<prp\_lxc1>...</prp\_lxc1>

Normal preposition

<prp\_lxc2>...</prp\_lxc2>

Dependent preposition

---

## ARTICLE

---

<at>...</at>

Article

---





---

## PRONOUN

---

<code>&lt;pn_inf&gt;...&lt;/pn_inf&gt;</code>	Pronoun inflection
<code>&lt;pn_agr&gt;...&lt;/pn_agr&gt;</code>	Number/sex disagreement of pronoun
<code>&lt;pn_cs&gt;...&lt;/pn_cs&gt;</code>	Pronoun case
<code>&lt;pn_lxc&gt;...&lt;/pn_lxc&gt;</code>	Lexis

---

## CONJUNCTION

---

<code>&lt;con_lxc&gt;...&lt;/con_lxc&gt;</code>	Lexis
---	-------

---

## RELATIVE PRONOUN

---

<code>&lt;rel_cs&gt;...&lt;/rel_cs&gt;</code>	Case of relative pronoun
<code>&lt;rel_lxc&gt;...&lt;/rel_lxc&gt;</code>	Lexis

---

## INTERROGATIVE

---

<code>&lt;itr_lxc&gt;...&lt;/itr_lxc&gt;</code>	Lexis
---	-------

---

## OTHERS

---

<code>&lt;o_je&gt;...&lt;/o_je&gt;</code>	Japanese English
<code>&lt;o_lxc&gt;...&lt;/o_lxc&gt;</code>	Collocation
<code>&lt;o_odr&gt;...&lt;/o_odr&gt;</code>	Misordering of words
<code>&lt;o_uk&gt;...&lt;/o_uk&gt;</code>	Unknown type errors
<code>&lt;o_uit&gt;...&lt;/o_uit&gt;</code>	Unintelligible utterance

---

# Examples

- ◆ There are <at crr= “a”></at> lot of books.
- ◆ There are <o\_lxc crr= “a lot of”>lot of</o\_lxc> books.
- ◆ They are <v\_lxc crr= “watching”>looking</v\_lxc> monkeys.
- ◆ They are <v\_lxc crr= “watching”>looking<prp\_lxc2 crr= “at”> </prp\_lxc2></v\_lxc> monkeys.

# สรุป

- ◆ คลังข้อมูลภาษาผู้เรียนเป็นทรัพยากรที่มีความสำคัญต่อการศึกษาวิจัยเรื่องการเรียนการสอนภาษาต่างประเทศเป็นอย่างมาก
- ◆ แม้ว่าการศึกษาเปรียบเทียบภาษาอังกฤษและภาษาไทยจะมีผู้ทำมาเป็นเวลานานแล้ว แต่ก็เป็นการเปรียบเทียบโดยไม่ได้อาศัยข้อมูลภาษาผู้เรียนเป็นหลัก
- ◆ การศึกษาข้อผิดพลาดที่ผ่านมาก็ทำให้เห็นเฉพาะข้อผิดพลาดแต่ไม่ได้เห็นถึงความสามารถของผู้เรียนว่ามีความเหมือนหรือต่างจากเจ้าของภาษามากน้อยเพียงใด
- ◆ การไม่มีข้อมูลกลางที่สามารถใช้ร่วมกันได้ ก็เป็นอุปสรรคต่อการศึกษาและหาข้อสรุปร่วมกัน การพัฒนาคลังข้อมูลผู้เรียนภาษาอังกฤษชาวไทยจึงมีความจำเป็นและจะเป็นประโยชน์ต่อการศึกษาวิจัยการเรียนการสอนภาษาอังกฤษในประเทศไทย