

**A Review of the Use of Learner-  
Behavior Tracking in CALL  
Research Studies:  
Programs Providing Meaning Aids**

**Fenfang Hwu  
University of Cincinnati**

**CALICO, University of  
Victoria, Canada  
May 19, 2011**



# Introduction

- Using tracking technology to gain insights into the way students learn or process language information can be traced as far back as to the 1980s. Nevertheless, researchers continue to face challenges in collecting and studying tracking data.



# Objective of the Study

To suggest tracking data sharing through data repositories as a way to accomplish the following goals:

1. To ease the challenges researchers face in collecting and studying tracking data in non-CMC environments (e.g., reading, listening applications)
2. To increase the use of tracking data to gain insights into second language learning
3. To synthesize and enhance CALL research

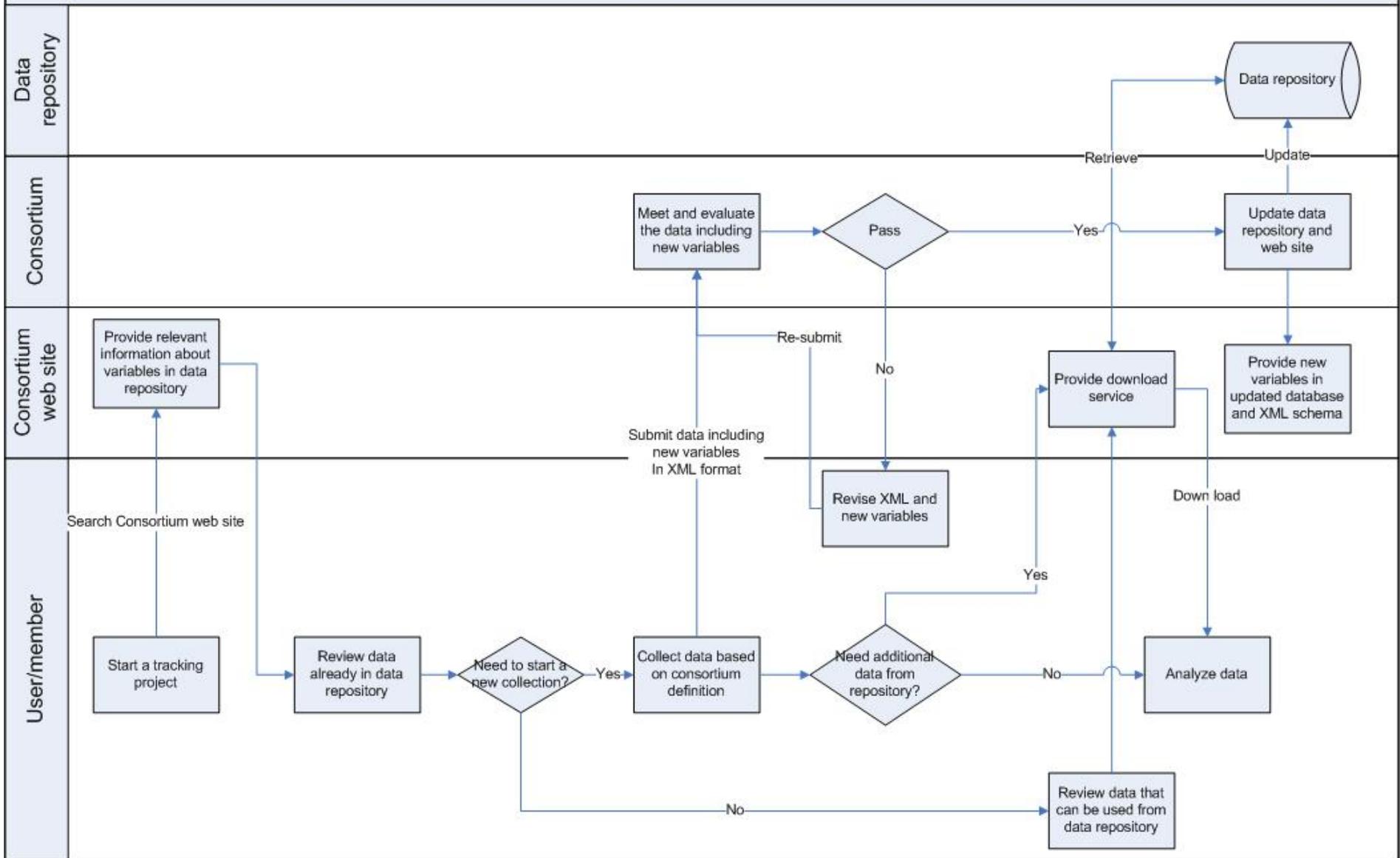
# What are Meaning Aids in CALL?

- Meaning aids are devices assisted by computer technologies with the purpose of providing the meaning of a language item during a learning activity (e.g., reading, listening) (Hubbard, 2000).
- They may include glossaries, dictionaries, summaries, explanations, translations, and audio/video scripts, which typically can be accessed instantaneously through the click of a button (Hubbard, 2000).

# This Presentation

- I will first outline the functions of a potential data repository
- I will then review extant research studies. The ultimate goal of the review is to identify
  - the research questions that can be investigated using the data from a data repository
  - the required data fields for the database table in the repository
- Finally, I will provide an example that shows why using datasets from the data repository can enhance the generalizability of CALL findings

# Process flow of using and contributing to data repository





# The Review

- Reviewed about 20 research papers, in which computer-based tracking was used to record second language learning behaviors in reading and listening applications.

# Research Qs Answered Using Tracking

Due to time constraint, the presentation will focus on the review of incidental vocabulary learning through online reading activities.

- (1) categorizing learning styles and gloss-usage types
- (2) effects of whether learning styles are accommodated on vocabulary retention
- (3) effects of *looking up* glosses on vocabulary retention

# Research Qs Answered Using Tracking

- (4) effects of the attributes of the instructional environment on the frequency of lookups
- (5) effects of the attributes of the instructional environment on the frequency of lookups and consequently the vocabulary retention
- (6) effects of the overall reading time on vocabulary retention
- (7) identifying vocabulary learning strategies

# Why Identifying These Qs?

- Q#3 can be re-analyzed using newly selected datasets, or a new statistical or theoretical approach (will address this in more detail).
- When addressing this question, previous studies used certain statistical approaches.
  - Some studies investigated the *correlation* between the total number of words consulted and the vocabulary test score. (e.g., Chun & Plass, 1996; knight, 1994)

# Why Identifying These Qs?

- Previous meta-analyses compared the performance (the effect size) of the *experimental* group (with glosses) with the *control* group (no glosses) (e.g., Abraham, 2008; Yun, 2011). In other words, they did not exclude studies that did not use tracking to confirm the consultation of a gloss. Many studies assumed that the availability of glosses always leads to their usage and consequently vocabulary retention.

# Why Identifying These Qs?

- Application design, learning style, time on task, or learning strategy seem to be important questions to CALL researchers. But they have not yet been addressed by research synthesis (e.g., meta-analytical studies).
- They may be answered by using the aggregated datasets from the data repository.

# Behavioral Data Collected by Tracking

- (1) Words readers clicked
- (2) Annotation properties associated with the words clicked (e.g., text, picture, dictionary, L1, L2, visible/invisible links, inferable/not inferable in meaning)
- (3) Time (hours, minutes, and seconds) when readers clicked a button or link, i.e., time stamps
- (4) The order in which readers selected words and annotations

# Behavioral Data Commonly Derived

Straightforward interpretations, compared to other types of applications (e.g., grammar, drills). However, derived data vary from study to study (will elaborate on this later)

- (1) the total number of queries learners made on a type of annotation
- (2) **the total number of queries learners made on annotations**
- (3) the total number of words of an individual type of annotations learners consulted
- (4) **the total number of words learners consulted**
- (5) the total amount of time learners spent on clicking annotations
- (6) the total amount of time learners spent on reading-related activities
- (7) the annotation type first consulted by learners
- (8) the number of words double-checked by learners.
- (9) the number of words or the percentage of words remembered by each gloss-usage or learner group.

# Variables Considered in Research Studies

- (1) Behavioral data collected by tracking (e.g., words readers clicked)
- (2) Performance data collected by hand (e.g., the score of multiple-choice or recall test) (although such data can also include the answers learners gave during practice and can be collected by tracking)

# Variables Considered in Research Studies

(3) Attributes of the instructional environment, collected by tracking or other methods; the list is extensive

- a. learners' characteristics (e.g., gender, age range, instructional level, L1, L2, learning styles, working memory)
- b. nature of learning tasks (e.g., whether words were relevant for reading comprehension questions, whether learners were forewarned of an upcoming vocab. test)
- c. media characteristics (e.g., marked vs. unmarked annotation condition; expository vs. narrative)
- d. assessment types (e.g., receptive vs. productive) (Abraham, 2008)



# The Data Fields Required for the Repository

## (1) Behavioral data

The repository consortium will suggest that all the behavioral data that can be collected by tracking be collected, which can be done without any added effort.

## (2) Performance data

The repository consortium will request that the data submitted contain at the minimum one type of behavioral data and one type of performance data to ensure that these data can be used by others.

# The Data Fields Required for the Repository

## (3) Attributes of the instructional environment

The repository administrators will recognize that the range of variables submitted may vary from study to study.

(4) To ensure that the data submitted can be shared among researchers, the repository administrators will require that such data contain certain information, such as the researcher's ID, the subject's ID (to maintain the anonymity of data), the subject's L1 and L2, and the target words.

# Suggestions to Increase the Generalizability of CALL

- 1) Share CALL applications (Levy, 2000)
- 2) Increase the number of research participants and/or the length of studies (Hubbard, 2005)
- 3) Conduct various modes of research synthesis
  - meta-analysis (e.g., Abraham, 2008; Taylor, 2006; Yun, 2011)
  - narrative or systematic review (e.g., Dunkel, 1991; Felix, 2005, 2008)



# **Suggestions to Increase the Generalizability of CALL**

- 4) Conduct secondary data analysis and replicated secondary data analysis



# What is Secondary Data?

- Data collected by someone other than the original user.

# Values of Secondary Data

- Reanalyze existing datasets with novel statistical or theoretical approaches (Smith, 2008, p. 3; NIH)
- Examine how a different analytic approach might change the conclusions reached from the first analysis (Boslaugh, 2007, p. 2).
- Address questions not considered in the original analysis (Boslaugh, 2007, p. 2; NIH).

# Values of Secondary Data

- Explore new research questions (Smith, 2008)
- Provide Insights into problem during exploratory phase
- Replication and confirmation of results (ICPSR)
- An efficient way of bringing together a large amount and breadth of data (Boslaugh, 2007; NIH)

# Values of Secondary Data

- Ensures that resources spent on data collection are put to the best use possible (ICPSR)
- Teaching (ICPSR; NIH)
- Unpublished data can be included

# How Widely Are Secondary Data Analyses Used?

- Smith (2008) analyzed published articles in eight well-regarded social science and education journals over a 7-year period and estimated that close to 10% of papers involved the analysis of secondary data.

# What Need to be Considered in Secondary Data Analysis?

- Like meta-analysis (e.g., Abraham, 2008; Taylor, 2006; Yun, 2011)
  - It may need to select the datasets to be included according to certain criteria (e.g., comparable studies)
  - It may need to designate some variables as moderator variables (e.g., proficiency level, assessment types)
- Unlike meta-analysis, it may need to convert datasets into the appropriate format



# An Example of Using Data from Repository

As mentioned earlier, Q#3 can be re-examined using newly selected datasets, or a new statistical or theoretical approach

#(3) effects of *looking up* glosses on vocabulary retention



# Re-Analyzing Existing Data Using New Approaches

The theoretical approach:

- The noticing hypothesis (Schmidt, 1990)
- Use computer's tracking to provide a concurrent measurement of noticing of *vocabulary* (Hegelheimer & Chapelle, 2000)

# Re-Analyzing Existing Data Using New Approaches

## Definition of “noticed”:

- Gloss clicked = noticed = 1
- Gloss not clicked = not noticed = 0

For each of the target word, two values are obtained:

- Noticed, remembered = 1, 1
- Noticed, not remembered = 1, 0
- Not noticed, not remembered = 0, 0
- Not noticed, “remembered” = 0, 1

# Re-Analyzing Existing Data Using New Approaches

- The statistical approach:
  - a phi correlation calculated for *each learner* to summarize the relationship between the words noticed and those retained (Hegelheimer & Chapelle, 2000).
  - The unit of analysis is the *word* rather than the *learner* (Hegelheimer & Chapelle, 2000).

# What Need to be Done to the Existing Data?

- Data values need to be converted  
1 = noticed      0 = not noticed
- A phi correlation need to be calculated; using the new number assigned to each case and the corresponding retention score

## Two Sets of Hypothetical data inspired by previous studies

Researcher ID	Subject ID	L1	L2	Target word	Frequency L1 verbal gloss clicked	Frequency L2 verbal gloss clicked	Frequency visual gloss clicked	Retained?
R0001	S0001	Eng	Germ	FischschwErme	1		0	1
R0001	S0001	Eng	Germ	Hubschrauber	1		1	1
R0001	S0001	Eng	Germ	Zigarette	0		0	0
R0001	S0002	Eng	Germ	FischschwErme	2		0	1
R0001	S0002	Eng	Germ	Hubschrauber	1		1	1
R0001	S0002	Eng	Germ	Zigarette	0		1	0
R0002	S0001	Span	Eng	launching		1		1
R0002	S0001	Span	Eng	era		1		0
R0002	S0001	Span	Eng	conceive		0		1
R0002	S0002	Chin	Eng	launching		1		1
R0002	S0002	Chin	Eng	era		1		0
R0002	S0002	Chin	Eng	conceive		2		1



# Summary

Q: Why reviewing CALL research studies?

A: To identify the data fields required for the data repository and the research questions that can be investigated using secondary data.

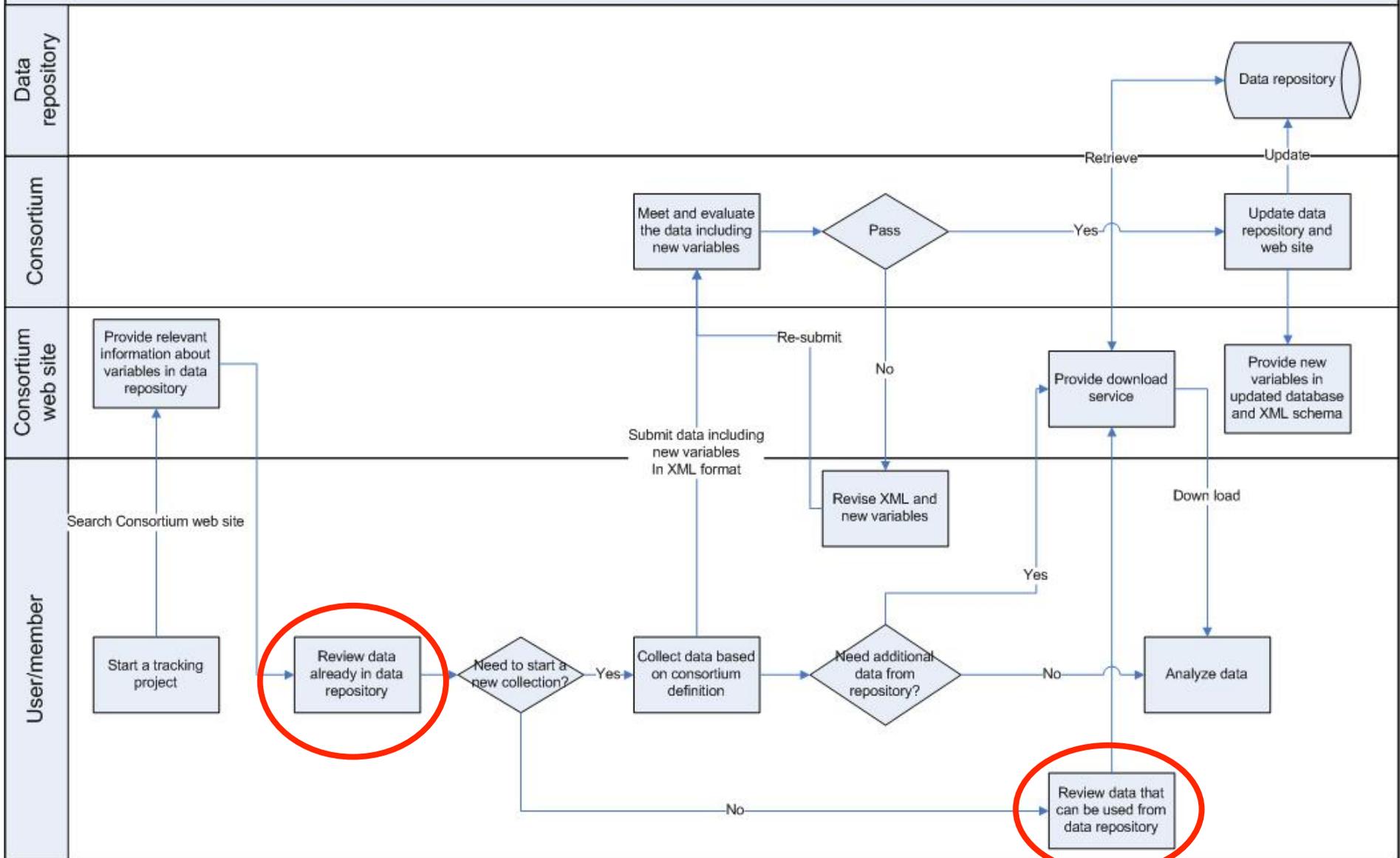


# Summary

Q: How does a data repository ease the challenges researchers face in collecting and studying tracking data?

A: Researchers can find out what variables have been examined by researchers, examine existing data, and obtain secondary data from the repository (See the graphic next page).

# Process flow of using and contributing to data repository





# Summary

Q: Why can a data repository enhance CALL generalizability?

A: It provides data, which can be re-analyzed using novel statistical or theoretical approaches (NIH; Smith, 2008) and brings together a large amount of data (Boslaugh, 2007; NIH).

# Data Repository Examples

UC Irvine Machine Learning Repository

National Center for Ecological Analysis and Synthesis (NCEAS)  
Data Repository

Geological Society of America (GSA) Data Repository

Open Language Archives Community (OLAC)

Child Language Data Exchange System

TalkBank

NIH supports the concept of data sharing, expects the timely release and sharing of data of grant receivers, and has provided guidelines on this issue.

# Conclusions

- To facilitate data sharing and advance CALL research, there is a need to establish a standardized way to describe data (e.g., proficiency level, level of instruction).
- Because the derived behavioral data vary from study to study (e.g., some looked at the total number of *words* clicked while others looked at the total number of *clicks* made), there seems to be a need to establish a standardized way to analyze these behaviors.

# Conclusions

- Although research about meaning technologies is probably the second most frequently investigated CALL study (following CMC), more research studies need to use tracking technology to investigate the effects of meaning aids in *listening comprehension* applications.

# Conclusions

- Because the repository will eventually contain a large amount of data and a large number of variables, it is anticipated that researchers will be able to analyze these data using data mining applications. Consequently, new insights into student learning processes may be discovered.

# Thank you...

## Contact Information:

Author: Fenfang Hwu  
Associate Professor  
Dep. of Romance Languages & Literatures  
University of Cincinnati  
fenfang.hwu@uc.edu

Co-author: Ching-Yeh Tzseng  
Enterprise Application Architect  
ctzseng@yahoo.com