Why use Concordancers?
- Possible benefits:
  - easy access to “real” language use
  - increase analytic learning
  - promote explicit L2 knowledge
  - gain critical language awareness
  - increase learner autonomy
  - sensitize learners to differences between their L1 and L2, as well as between different genres of the L2 (this would aid students in selecting appropriate language in a given situation)

Theoretical Background
- Rationale behind concordancing as a teaching tool is that:
  - we learn languages by forming and testing hypotheses about a language
  - this is usually done subconsciously
  - if we make this a conscious process, we will learn better
- Problems with this rationale:
  - we can only make hypotheses about language differences that we actually notice
  - if the differences between the languages are subtle, they are easier to miss
  - because language differences between advanced ESL learners and native English speakers are much more subtle, it is harder for an advanced learner to notice differences

Drawbacks of Available Concordancing Software
- need to have a hypothesis about a language feature before you search
- not very useful in hypothesis formation
- students will need guidance in what features to look for (not good for autonomous learning)
- too much data can be overwhelming and frustrating

Corpus-Linguistic Study and Analysis
Method: Comparison between native and ESL-learner corpora and analysis of the results for over-indulgence and under-representation to determine if this learning method is efficient.
Data Used: Intercultural email projects between advanced German ESL learners and American native English speakers were submitted as the basis for the corpora.
Results: The ESL learners either over- or under-used more than 1,000 words significantly when compared to their native counterparts.
Conclusions:
- the contexts in which a target word appears must be analyzed
- equivalent words to the target (ie.synonyms) must often be included in the analysis
- the relationship between the target and its equivalent words is what will lead to the formation of hypotheses by the learner. These hypotheses can then be tested.
- there are many causes behind over-indulgence and under-representation of a particular target word:
  - use of different standard varieties of the L2
  - preference of certain subjects (caused by differing interests or cultures)
  - interlanguage idiosyncracies in the L2 corpus
**Some Definitions**

**Interlanguage:** a language system created by learners of a language, composed of elements from both the native and the target language. The learners themselves give structure to the linguistic data they have and create this internalized system (the interlanguage). It is comprised of many stages, corresponding to each level of the target language learning (Gass & Selinker 1994, 11).

**Concordancer:** “a tool for text analysis which can generate lists of the words contained in a text or text collection (corpus)” (Gabel 2001)

**Over-Indulgence:** excessive use of a particular feature of an L2 (for example, certain lexical items or grammatical structures)

**Under-Representation:** absence or the infrequent use of a particular L2 feature (for example, certain lexical items or grammatical structures)

**Corpus:** the body of text through which the software searches for target words or phrases.

**KWIC:** Key-Word In Context. User enters a word or phrase, the program searches through it’s database and gives a list of all the contexts in which this word or phrase occurs.

**Frequency List:** Within a specific text (or texts) the user enters a word or phrase, and the number of times (frequency) this word occurs is given as the output. Problem: gives too much data.

**Contrastive Frequency List:** Compares text(s) of different language types (ie. native and non-native) looking for a specific word or phrase and counts the number of times this word or phrase occurs in each text (results are adjusted for size of each text). Problem: doesn’t say if the language differences are significant.