

Designing Concordance-Based Vocabulary Activities¹

Jesús del Carmen Manjarrez Moreno², Universidad Da Vinci, Mexico City, Mexico

Abstract

This paper is focused on one of the most common and basic ways of processing corpus information; concordance, and how it can be used in the classroom to offer language learners useful vocabulary that they would face in genuine conversations and help them detect language patterns as this ability helps them with their learning process. In addition, a sample of a corpus-based activity will be presented to show how classroom activities based on corpora can be developed. What is more, this paper has the purpose of encouraging teachers to use this type of activity and to offer them a straightforward guide to follow.

Resumen

Este documento está enfocado en el método más común y básico para procesar la información de un corpus; concordancia, y cómo esta información puede usarse en el salón de clase para ofrecer a los estudiantes de idiomas vocabulario útil, que podrían enfrentar en una conversación real, y ayudarlos a desarrollar la habilidad de detectar patrones en el lenguaje, ya que esta habilidad les beneficia con su proceso de aprendizaje. Además, se incluye una actividad basada en información de corpus para demostrar el proceso que puede seguirse al crear una actividad de clase, permitiendo así, analizar la utilidad de un corpus. Sobre todo, este documento tiene el propósito de alentar a los maestros a usar este tipo de actividad y les ofrece una guía sencilla a seguir.

Introduction

Corpus might have a very general definition as the one given by The Experts Advisory Group on Language Engineering Standards (EAGLES) mentioned by Meyer (2004), which states that corpus might be any text type collection as newspaper, poetry, drama, dictionaries, etc. Perhaps, linguists analysing corpus prefer a more restricted definition, referring to a collection of text used for linguistic analysis. Collins COBUILD English Language Dictionary, published in 1987, was the first corpus-based dictionary and the one that brought corpora alive for English teachers (Grabrielatos, 2005). Since then, corpus-based products have grown exponentially, not only teaching activities, products related to this area as dictionaries or frequent word lists, but also linguistic products as well.

Besides, some different corpora definitions there exist different corpora. They differ in shapes and size because they are built for different purposes. "There are two philosophies behind their design, leading to the distinction between *reference* and *monitor* corpora...Another design-related distinction is whether a corpus contain *whole texts*, or merely *samples* of a specified length" (Gabrielatos, 2005, p. 3).

- Reference corpora: It has a fixed size (e.g., The British National Corpus).
- Monitor corpora: It is expandable (e.g., The Bank of English)

In terms of content, corpora can be *general* or *specialized*:

- General corpora: Reflect a specific language or variety in all its contexts. (e.g., the American National Corpus)
- Specialized corpora: Focus on specific contexts and users (e.g., Michigan Corpus of Academic Spoken English)

Corpora is relevant to language teaching in many ways since concordance offers realism and relevance. It is a common practice for language teachers to create inauthentic activities which are correct in the target language, but might not exist in a real context. Concordance-based materials can offer real and relevant texts (Ma, 1993). In addition, "corpora of language teaching coursebooks enable the examination of the language to which learners are exposed" (Gabrielatos, 2005, p. 4).

The importance of a language corpus is that it contains examples of realm up-to-date language use. In the past, researchers used the language they heard or read from people in their circle of friends as examples. These examples were intuitive and they were often correct or nearly correct, as stated by Fox (1998), but they were not always reliable or related to what people, in general, were really saying. Currently, teachers still use introspection to create their activities and while many grammar structures are simplified and some exceptions are presented, many other uses of the language are ignored.

¹ Received: 20 May, 2021. Accepted: 11 September, 2021. Published: 15 February, 2022.

² lj.c.manjarrez@hotmail.com, contacto@udavinci.edu.mx

Corpus-based research has also revealed that many common language patterns are missing in ELT materials. Corpora help to analyse the language forms that are used by native speakers in which context and frequency give language learners a more reliable model to follow. For example,

The transitive use of 'give' in clauses such as 'she gave him a really lovely smile', and 'he gave an extremely boring talk' is so common that many native speakers hardly even notice it. The meaning of 'give' here is easily intelligible to learners, but what they might not realize is how important this particular structure pattern is...it allows us to focus on the event rather on the action (Fox, 1998, p. 27).

Concordance Method for Corpora Searching

According to Hunston (2002), "Producing concordance lines is the most basic way of processing corpus information" (p. 38). Concordance lines are "all the instances of a word or phrase in the corpus." (Bennet, 2010, p. 17) To obtain concordance lines from a corpus, a concordancer is needed. This is a program that searches a word or phrase in a corpus and then presents it on the computer screen with the words that come before and after it. The word that is searched is called node word.

The concordance lines produced in the search can be sorted in alphabetical order using the word to the left or the right side of the node word for the sorting. This sorting makes it easy to analyse the results and find patterns on how that node word is used (See Appendix 1). Understanding concordance lines needs some practice and as mentioned by Breyer (2009), it is important that teachers receive corpora and concordance teaching during their initial training programs. She expressed that since corpora is still not included in the curricula, it is the teachers' decision to incorporate corpora-based activities in their teaching at present time.

Breyer (2009) pointed out some key factors that influence teachers' decision on implementing corpora into their teaching:

- Motivation.
- Availability of materials.
- Possession of adequate skills to teach with corpora.

Throughout her investigation, Breyer (2009) concluded that the best moment to include corpus training was during initial teacher education; "in this context, the potential exists to integrate it meaningfully into their training and to devote sufficient time to the subject to allow for in-depth exploration" (p. 4). It is important though, to teach the teacher learners their role as teachers when applying corpora in their classes, this is not the typical role a teacher has. When using corpora, the teacher "has to learn to become a director and coordinator of student-initiated research" (Johns, as cited by Breyer, 2009. p. 4).

The results presented in Breyer's (2009) case study in which student teachers initiated their corpora training, it can be observed that the student teachers were concerned at the beginning about how language teaching had become so simplistic while trying to make language learning easier for learners. They felt that process was not giving learners the correct picture, as demonstrated in the following comments:

I suppose it could be a problem if they just learn the general rules and then, eventually, are confronted with sentences which do not fit into the system they were taught. I assume this is very confusing and it is hard to look beyond the rules, which one has once learned. (Breyer, 2009, p. 10).

The point is, however, that over-simplification leads to incorrect portraying of authentic language use...Pupils are taught only a part of the rule in a bid to keep it straightforward and simple (Breyer, 2009. p. 10).

Later, while student teachers were advancing in their corpora training, they came up with some ideas on how to incorporate the use of corpora. As stated by Breyer (2009), student teachers suggested the following procedure:

- Teach the basic language rules for beginners.
- Once language learners had learned basic language rules, they can discover extended rules on their own using a concordancer. Data-driven learning is a good option.

This study gave a clear view about how student teachers became aware of the importance and usefulness of corpora in their teaching, and demonstrated it could be useful to introduce corpora training into teaching education.

What to observe in concordance results

Once teachers have been trained on how to use corpora, they will be able to make the correct observations; there are different types of observations that help to obtain useful information from concordance:

1. Observing the *central* and *typical*.
2. Observing *meaning distinction*.
3. Observing *meaning and pattern*.
4. Observing *detail*.

Typical: "Might be used to describe the most frequent meaning or collocates or phraseology of an individual word or phrase" (Hunston, 2002, p. 42). For example, if you search for the concordance lines for the word "recipe" you would be able to observe that the *typical* meaning for this word is metaphoric, not a literal recipe, as in the following concordance line:

"to god material, and you have a recipe for serious success. But is not a" (2002, p. 43).

Centrality: "Can be applied to categories of things rather than individual words" (Hunston, 2002, p. 43). For example, talking about the uses given to the present progressive; it is known that it is used to talk about the present, the future, or not specific time; however, it is more frequently used to talk about the present; therefore, she stated that the central use for present progressive is to talk about the present time.

Among these two concepts, surges a third one: *prototypical*; it is used to indicate a common usage that can be perceived as typical, but is not necessarily the most frequent. It has been demonstrated that prototypical examples are frequently used in language coursebooks (Hunston, 2002). An explicit example is the way that *comparatives* are presented, where the grammar form indicates that the word *than* is part of a comparative sentence, but concordance has demonstrated that typically it is not used as in the following sentences: "a much larger plan, or their larger but poorer northern neighbours" (p. 44). In these previous sentences, the comparison is implicit.

Meaning distinction: There are words with similar meaning, but they cannot replace one another; they are called near-synonyms. Concordance lines can help to clarify their different meanings.

Meaning and pattern: "Meaning of a word is closely related with its co-text" (Hunston, 2002, p. 46). Its meaning depends on the patterns or phraseologies where they occur. If the pattern is distinguished, then the meanings are as well. This applies when searching for individual words. There is also the option to obtain meaning for words with similar meaning that share patterns. This last option can be done by "searching for phrases and noting the words that frequently occur within it" (p. 48).

A variation of searching phrases to obtain meaning is the concept of frames; It is a sequence of, usually three words in which the first and last are fixed, but the middle word is not. For example:

too + adjective + to → too young to, too easy to, etc.

Detail: Observing details helps to lead to more specific observations about the specific behaviour of individual words.

Concordance in the Language Classroom

As mentioned before, concordance lines can be exploited in the language classroom to help make language learners aware of how language is used. This has been demonstrated by previous studies which show how the concordance approach to teaching vocabulary is more effective than traditional methods, as demonstrated by Gan et al. (1996).

In their study, Kaur and Hegelheimer (2005) analysed the effectiveness of using an online concordance and dictionary in acquiring new vocabulary, and comparing this procedure with using only an online dictionary. The study focused on helping learners to improve their academic writing skill by increasing their vocabulary. This study highlighted the importance of having extensive vocabulary knowledge for ESL learners in order to produce the types of writing expected from them in an academic setting, and re-marked how difficult it is for ESL learners to acquire vocabulary due to the limited direct contact they have with the target language. Therefore, ESL learners are limited to produce language due to their poor lexical capability. In this study, Kaur and Hegelheimer, worked with a group of learners who were enrolled in a writing course for ESL undergraduates after taking the English Placement Test (EPT). Their writing evaluation showed deficiency in their composition writing. A group of language professors selected thirty words from the Academic Word List (Coxhead, 2000) for the learners to use to develop a specific writing task. The concordancer selected was Tom Cobbs's Compleat Lexical Tutor and the Dictionary.com was the online dictionary used. Learners were indistinctly separated into two groups; one group could use the concordancer and online dictionary and the other one could not.

The learners went through the following procedure:

1. Answering a questionnaire to identify their language.
2. Exposure to the use of a concordance program.
3. Answering a pre-test to evaluate their receptive academic vocabulary knowledge.
4. Solving a cloze activity.
5. Solving a sentence-building task for learners to demonstrate their understanding of the meaning.
6. Trying to collocate the target vocabulary to observe if they could use it productively.
7. Solving a writing task where learners could demonstrate their productive knowledge.
8. Answering a post-questionnaire to obtain opinions from learners about how positive the experience with concordance was.

After this procedure, it was demonstrated that those learners who had access to the concordancer and online dictionary had improved their performance on using the target words correctly than those who did not have access to these tools. In addition, it was also showed that both groups of learners used the target vocabulary in their writing task. Nevertheless, the ones working with the concordancer and online dictionary made more attempts and had more correct words compared to the learners who did not use these tools.

Therefore, it is useful to encourage language teachers to develop corpus-designed activities, to help learners interact with native-like language use and learn from this interaction.

A framework for creating corpus-designed activities

Direct use of corpora for teaching includes three different focuses according to Leech (1997):

- Teaching about: Teaching corpus linguistic as an academic subject.
- Teaching to exploit: Teaching learners how to use corpora by themselves (hands-on). The learning activity may become learner-centred.
- Exploiting to teach: Using a corpus-based approach to teaching language and linguistics courses.

As McEnery and Xiao (2011) expressed, teaching about and exploiting to teach are associated with students of linguistics and language programs, while teaching to exploit is related to students of all subjects including language study and learning who are expected to derive benefit from Data-driven learning (DDL) and Discovery Learning.

There are different approaches to creating corpus-designed activities like the three Is approach (Illustration, Interaction, and Induction) mentioned by McEnery and Xiao (2011), where illustration refers to looking at real data, interaction is discussing and sharing opinions and observations, finally, induction refers to making one's own rule for a particular feature of the language. This approach has the drawback of provoking partial and incomplete generalizations from limited data "as a stage on the way towards a satisfactory rule" (p. 370).

Another approach is the Data-driven Learning (DDL) suggested by Johns (1991). This approach is learner-centered, and guides the learners in an autonomous learning process where the learner becomes the researcher. Since making a correct interpretation of corpora results is not an easy task, McEnery and Xiao (2011) mentioned that teachers' guidance and pedagogical mediation is needed. Therefore, as mentioned before, corpus training for language teachers is necessary.

The DDL (Data Driven Learning) approach includes three stages of inductive reasoning: observation of concordance evidence, classification of salient features, and generalizations of rules, similar to the 'Three Is' approach (Johns, 1991).

Bennet (2010) suggested the following framework which offers seven clear and easy-to-follow steps. The first five steps are for the teacher to identify the target language to work with, and the type of corpus to be used. The last two steps are related to the activity creation where the 'three Is' or 'DDL' approach might be used.

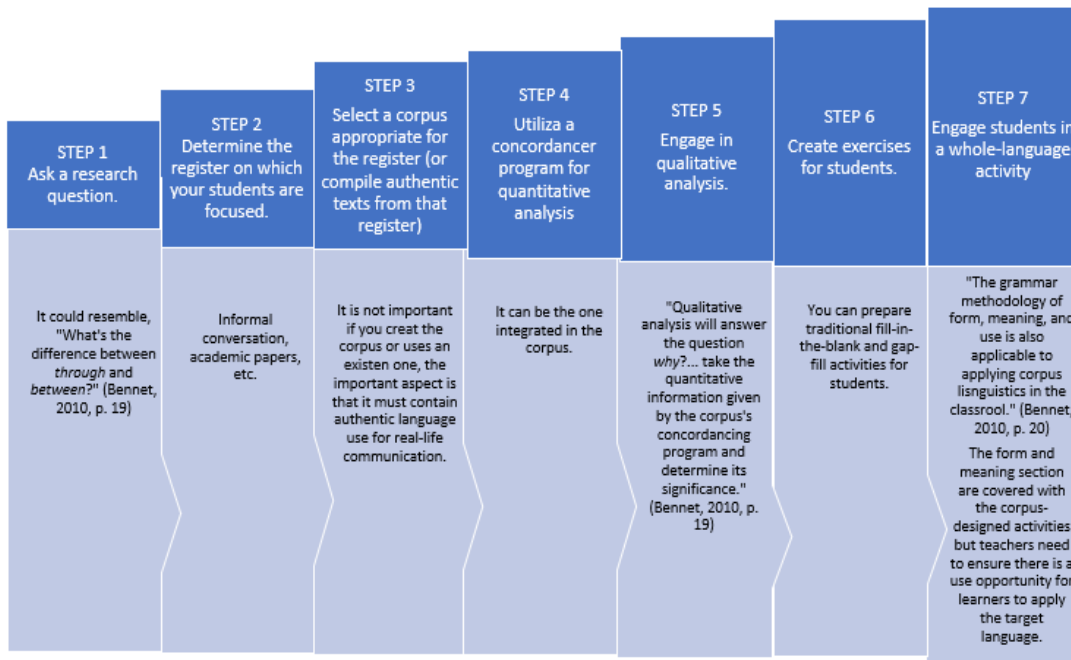


Figure 1. A framework for developing corpus-designed activities (Adapted from Bennet, 2010, p. 19)

A similar framework, developed in six steps, is mentioned by Bardovi-Harlig et al. (2015):

1. Selecting the corpus
2. Identifying expressions
3. Extracting examples
4. Preparing corpus excerpts for teaching
5. Developing noticing activities
6. Developing production activities

This framework was used to analyse expressions since the focus of the article was on pragmatic routines (expressions), such as *Thank you so much* and *That'd be great!* for thanking and *Nice to meet you* for introductions. It can be observed, in this second framework that the corpus is selected before identifying the expression (*research question* step in Bennet's (2010) framework). Once these two steps covered the following steps, even though expressed differently, are similar in function: They use the selected corpora, analyse the results, adapt the aspects to facilitate the target language (TL) coverage, create the corpus-based activities for learners to notice the TL and then provide learners with the opportunity to experience/produce it.

Application in a specific context

As an example, this paper presents an activity created for intermediate level students in a language school, which belongs to a public university in Mexico. These students are in their final semester of a three-year-long English program.

This activity was designed to clarify the use of the word *miss*, since the students seemed to be confused with its use and meanings. The activity development follows Bennet's framework. It was selected since it guides the researcher in a detailed form through the entire process; the steps to follow present a logical sequence.

Bennet's (2010) Framework A Framework for Creating Corpus-Designed Activities	
• Ask a research question	<i>What is the correct use for the word miss?</i>
• Determine the register on which your students are focused.	<i>Informal</i>
• Select a corpus appropriate for the register:	<i>COCA (Corpus of Contemporary American English)</i>
• Utilize a concordancing program for quantitative analysis.	The concordancing program used is the one integrated in the corpus.

<p>The results obtained were around 100 concordance lines. They were sort in alphabetical order based on the word to the left of the keyword miss.</p>
<ul style="list-style-type: none"> • Engage in qualitative analysis. <ol style="list-style-type: none"> 1. The first observation was based on the results of a KWIC search, guided by the colours, identify types of words, that the word miss can be used as <i>noun</i> and as a <i>verb</i>. “, but added, “It would be a sin to miss this opportunity” (verb) “Hold on one moment, please- Hi, Miss Acosta? -Yes- I 'm Dr. Lopez.” (noun) 2. The second observation is that most of the time, it is followed by a <i>proper noun</i> or by <i>pronoun</i>. 3. The third observation is that when used as a verb, the word miss has two different meanings: <ul style="list-style-type: none"> • To lose something/someone. “, but added, “It would be a sin to miss this opportunity” (verb) • To regret the absence of someone/something “I 'm still in love with Stacy, and I miss her.” <p>Note: The total steps and graphic results to obtain these results will be analysed in Appendix 1.</p>
<ul style="list-style-type: none"> • Create exercises for students. <p>From the 100 concordance lines, a group of 19 lines were selected to create a grammar awareness activity.</p> <ul style="list-style-type: none"> * Seven using MISS as a verb, meaning to lose something/someone. * Seven using MISS as a verb, meaning to regret the absence of someone/something. * Five using MISS as a noun. <p>The activity gives information to students about the correct use for the word MISS, and asks them to identify this usage in a list of concordancing lines. The complete activity will be analysed in Appendix 2.</p>
<ul style="list-style-type: none"> • Engage students in a whole-language activity. <p>The students will do the following writing task as homework as an opportunity to apply the knowledge. Imagine you are traveling with your family, and you are far from your best friend/girlfriend/boyfriend. Write this person a message to share how you are doing. Remember to include the word MISS in its three different meanings.</p>

Figure 2. Process to create a corpus-designed activity

Conclusion

Corpus-based activities give students contact with real use of language, which prepares them for real-world interaction. These activities could be as varied as the teacher's imagination is. Students learn to observe patterns, and it develops their language awareness helping them to be independent learners which increases the probabilities of their success in their language use. There is one important aspect to consider; teachers need to be trained in corpus use, searching techniques and corpus-based activities creation since their first encounter with corpora might be confusing.

References

- Bardovi-Harlig, K., Mossman, S., & Vellenga, H. E. (2015). Developing corpus-based material to teach pragmatic routines. *TESOL Journal*, 6(3), 499-526. <https://doi.org/10.1002/tesj.177>
- Bennet, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. University of Michigan Press.
- Breyer, Y. A. (2009). Learning and teaching with corpora: Reflection by student teachers. *Computer Assisted Language Learning*, 22(2), 153-172. <https://doi.org/10.1080/09588220902778328>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. <https://doi.org/10.2307/3587951>
- Fox, G. (1998). *Using corpus data in the classroom*. Cambridge University Press.
- Gabrielatos, C. (2005). Corpora and language teaching: just a fling or wedding bells? *TESL EJ* 8(4), 1-35. <https://www.tesl-ej.org/wordpress/issues/volume8/ej32/ej32a1>
- Gan, S.-L., Low, F., & Yaakub, N. F. (1996). Modeling teaching with a computer-based concordancer in a TESL preservice teacher education program. *Journal of Computing in Teacher Education*, 12(4), 28-32. <https://doi.org/10.1080/10402454.1996.10784301>
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research*, 1-16.
- Kaur, J., & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*, 18(4), 287-310. <https://doi.org/10.1080/09588220500280412>
- Leech, G. (1997). *Teaching and language corpora: A convergence*. Routledge.
- Ma, B. K. C. (1993). Small-corpora concordancing in ESL teaching and learning. *Hong Kong Papers in Linguistics and Language Teaching*, 16, 11-30. <https://files.eric.ed.gov/fulltext/ED365119.pdf>
- McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 364-380). Routledge.

Appendix 1

Corpus-based Activity Example

1.- Go to <https://www.english-corpora.org/coca>

2. To obtain KWIC concordance lines, click on the search section (1), select KWIC option (2), write in the searching box (3) the word you want to analyze, and click keyword in context (KWIC) (4) to start.

Note: Decide how many words to the left or to the right to use to order the results. Here one word to the left and to the right (5) were used.



3. The results will be displayed on the screen, and you can do your analysis.

The screenshot shows the search results page for the word 'MISS'. The results are displayed in a table with columns for rank, year, genre, source, and concordance lines. The concordance lines are color-coded to highlight the word and its context.

Rank	Year	Genre	Source	Concordance Lines
1	2004	NEWS	Atlanta	# * Oct. 23 Kentucky 1-8 # * Oct. 30 Die Miss B-S ; Photo : Georgia 's Uga V snaps at Auburn receiver
2	2019	MAG	Bleacher Report	line-up, played very well against Arsenal , now set to miss B-S months of action . # The 23-year-old has not quite kicked
3	2014	SPOK	CNN: Showbiz Tonight) (END-VIDEO-CLIP) HAMMER# Pretty incredible . She does n't miss a note . And Annalise says that she has n't had a
4	2019	NEWS	Columbus Dispatch	paying the price . # Vic Theado , Westerville Never miss a story # Choose the plan that 's right for you .
5	2019	SPOK	NPR_ATC	did n't know to look for it , you 'd probably miss a granite sign that reads Tucker Cemetery , first black family ,
6	1990	MAG	AccentLiving	need moral support before the next meeting .) If you miss a meeting or two , does someone call or drop you a
7	1998	SPOK	NBC_Dateline	: Did you say anything to the kids at school ? Miss A HENSEL : Hi . STAFFORD : You said , ' Hi
8	2005	MOV	Sueño	. Hold on one moment , please . - Hi Miss Acosta ? - Yes , - I 'm Dr. Lopez .
9	2013	SPOK	ABC: 20/20	women competing this year to become Miss Alaska . Backstage Miss Alaska 2012 is giving some pointers ... DEBBE EBBEN (MISS AL
10	2012	MOV	Shiver	to say . Well you 'd better think of something . Miss Alden , Miss Wendy Alden . Okay . Okay , I 'll
11	2018	TV	iZombie	dance . I 'm afraid if I dance , I miss Allan coming in . Or I will vomit . Okay . I
12	2000	MOV	Mission: Impossible II	souca na na **** Come on , come on miss allez , allez *** With a souca souca mama ****
13	2018	MAG	Vanity Fair	pageant writer Lewis Friedman . # The initial statement from Miss America 's board about the swimsuit decision claims it was "r
14	2012	WEB	weeklystandard.com	party offers . # Republicans never miss an opportunity to miss an opportunity , in 2010 , they failed to win the Senate
15	2014	TV	The Good Wife	n't . What if it 's important ? Then you miss an important call . Why were you giving me the silent treatme

Appendix 2

Classroom Activity

Activity 1

The word *MISS* can be used as a noun or as a verb.

Read the following sentences and write next to each of them an N if the word is being used as a noun, or a V if it is being used as a verb.

- 1.- n't. What if it's important? Then, you'll miss an important call. Why were you giving me the silent treatment?
- 2.- override! 792, shut down. Now! Did we miss anything important?
- 3.- buried under a mess of other issues. Whatever. I miss our friendship.
- 4.- her, that's all. Well, we'll all miss her. Yeah. It's nice to have a baby shower.
- 5.- Hold on one moment, please- Hi, Miss Acosta? -Yes- I 'm Dr. Lopez.
- 6.-world 's not holding its breath to find out what killed Miss Brown. She's a taxpayer.
- 7.- I 'm still in love with Stacy, and I miss her.
- 8.- wake up Miss Bundy? The dog ate my homework. Miss Bundy, why can't you be more like Ruben?
- 9.- dance. I 'm afraid if I dance, I'll miss Allan coming in. Or I will vomit. Okay.

Activity 2

The Word *MISS*, when used as a verb, has two meanings:

- To lose something/someone.
- To regret the absence of someone/something

Read the following sentences, and write next to each sentence an L if *MISS* means *TO LOSE SOMEONE/SOMETHING* or an R if it means *TO REGRET ABSENCE OF SOMEONE/SOMETHING*.

- 1.- he's selfish, that's why we don't miss him. And we've held it in for far too long.
- 2.- I run into Lessie, one of my roommates. "Miss me? "she says.
- 3.- sure and let me know. I don't want to miss it. Don't worry. You won't.
- 4.- There you are. I'm so sorry, miss Mara. My assistant put down the wrong terminal on my itinerary.
- 5.- First day didn't go so hot? [sighs] I miss my old school. Mm. Beginnings are tough.
- 6.- World. Right ... Daria, get moving or we'll miss the cruise. Just carry me to the car.
- 7.- "I'm not going to miss Dennis", Robinson says. "He brought a lot...
- 8.- but added, "It would be a sin to miss this opportunity.