



Identification, code-breaking and retention of idioms: relationships with idiom-type variables

Rosalina Domínguez Ángel
Universidad Autónoma de Tlaxcala

Abstract

This paper aims to explore the relationships found between idiom-type variables and product variables (i.e. immediate comprehension and delayed retention). Empirical investigation in the area of idiom comprehension and idiom retention has been focused on the exploration of some idiom variables such as: a) interlingual similarity, in the sense of how close in meaning the literal translation equivalent is (Irujo 1986, Cooper, 1999, Liontas 2002b, Charteris-Black 2002), b) level of formality, (Cooper 1999,), c) transparency (Cooper 1999) d) contextual inferencing (Cooper 1999, Liontas 2002) e) interlingual familiarity, similarity of metaphoric theme/vehicle (Boers & Demecheleer 2001, Charteris-Black 2002). One great disadvantage of the studies mentioned above is that scores of idiom comprehension and idiom retention have been correlated with one or two idiom-type variables but the analysis has not been done thoroughly, taking most of the possible idiom-type variables. Hence, in this paper we will report the results obtained in a study that analyses the relationship between idiom-type variables such as: L1-L2 interlingual similarity, transparency, contextual guessability and familiarity of words and indexes of comprehension and retention of unknown idioms in incidental conditions as the result of a reading task for Mexican EFL university students.

1. Introduction

In recent years, linguists, teacher trainers and language teachers have recognized the importance of formulaic sequences or multiword units including idioms, see (Irujo 1986a, 1986b, Howarth 1998, Bahns 1993, Nattinger & DeCarrico 1992, Lewis 1993, Alexander 1984, Cornell 1999, Lennon 1998, Cooper 1999, Yorio 1980, 1989. etc.). Idioms in the present study are those called 'phrasal idioms'. This means, phrases of more than only one word; conventionalized fixed or unfixed expressions with figurative unitary meanings that cannot always be determined from the meaning of their individual parts. Research on how formulaic



sequences and idioms are learnt is evidently lacking. The description of the development of learners' phraseological and idiomatic competence is essential in order to be better equipped for their teaching or for the promotion of the learner strategies that prove to be successful in identifying, comprehending and retaining idioms.

2. Previous research in the identification, comprehension and retention of idioms.

2.1 Idiom identification

The amount of research in idiom studies in L2 is limited and is even more restricted in the area of idiom identification. Concerning this area, there are two studies that have explored the identification of idioms: Liontas (2001) and Liontas (2002a) the first being the pilot study of the second. Both studies had as purpose the identification of strategies that learners used when searching for idiomatic phrases in authentic reading texts and measuring their success in identification. The participants in the main study ¹were fifty-three third-year university students of Spanish, French and German, being twenty-eight, twelve and thirteen students in each of the groups respectively asked to underline the words that constituted an idiom and they had to explain in a written report (retrospective protocol) why they thought the underlined words constituted an idiom. It should be noted that no figures for frequency of use of strategies are given. The strategy use in idiom identification was determined only by the number of idioms and percentages that learners were able to identify.

The results indicate that the percentage of correct idiom identification was of 86.33% including all groups of learners. The level of success of these subjects was much higher than in Liontas (2001) (70.09%). There was no significant difference across the groups concerning the success in identifying idioms. Nevertheless, the different idioms were detected more often than the other two types. The strategies that the learners used to identify the idiom were labelled according to the phrases

¹ This study is taken for the analysis because it deals with English idioms.



that the learners used in their reports as in Liantas (2001) and are the following: Literal meaning makes no sense, translation, contextual support, sounds /seems / looks like, prior knowledge and process of elimination. We see from the previous results that great attention has been paid to the use of identification strategies, but no relationships were searched concerning the type of idioms learners were dealing with and their success in identification.

2.2 Idiom Comprehension

Most of the studies on idiom comprehension and retention have been focused on one specific idiom-type variable that in the results appears as important for idiom comprehension and / or retention. One of the first studies focused on the comprehension and production of idioms was conducted by Irujo (1986). Her aim was to determine whether learners use their knowledge in L1 to help them understand and produce idioms in L2. The results showed that. Different idioms were more difficult to comprehend than identical and similar idioms (definition task 81% vs 96% and 97% respectively). Fewer different idioms were comprehended correctly and there was little evidence of interference for different idioms. In general the author concludes that advanced learners of a second language can use their knowledge of L1 to comprehend and produce idioms in L2 provided those two languages are related to one another. As we can see, the interlingual similarity of the idioms was central in Irujo's research.

A second study in the area of idiom comprehension was conducted by Cooper (1999). His aim was to explore the strategies that learners used to get the meaning of twenty idioms presented in a written context. The criteria to select the idioms was formality, this variable however did not show any important effect. The results showed that inferencing from context was the most successful of all the strategies hence, the amount of contextual clues within the text was another idiom-type variable explored.

Liantas' (2002b) tested the effect of context on the comprehension and interpretation of idioms in three different languages: Spanish, French and



German². The criterion for the selection of idioms was their interlingual similarity. The results in the context task reveal that context affects significantly the comprehension of idioms. Context favours the comprehension of different idioms more than identical or similar idioms which reached a high percentage right in the no context task.

2.3 Idiom retention

In the area of idiom retention, Boers and Demecheleer (2001), aimed to prove that idioms of medium transparency could be remembered better by learners with the help of guessing by using the imageability of conventional metaphors in L1 and L2. The variable involved in this study is the culture-specific grounding that certain idioms have (the 'vehicle' used for the metaphor). This study as well as Boers (2000), Bogaards (2001)³ and Charteris-Black (2002) show a special interest in metaphoric themes that could be shared by languages which are close one to the other. This is an additional contrastive L1 cognate clue kind of factor that in a way could be related to the variables investigated in Irujo (1986) and Cooper (1999) discussed above. The hypothesis that the researchers stated was apparently confirmed, concerning the success that students had in guessing the meaning of idioms whose metaphoric theme is close to their culture.

We have identified so far, within the three main areas of idiom research, idiom-type variables which are characterized by : a) interlingual similarity b) contextual clues c) metaphoric vehicle and d) transparency. This information structures the platform to do the design of our own study on the identification, comprehension and retention of idioms.

3. Design of the study

Our study aims to relate the idiom-type variables of 21 idioms contained in a reading text with learners' success in identifying, in code-breaking, and in retaining the meaning of the unknown target idioms after a three-week period. The idiom-

² The analysis of the data did not exclude other types of inferencing.

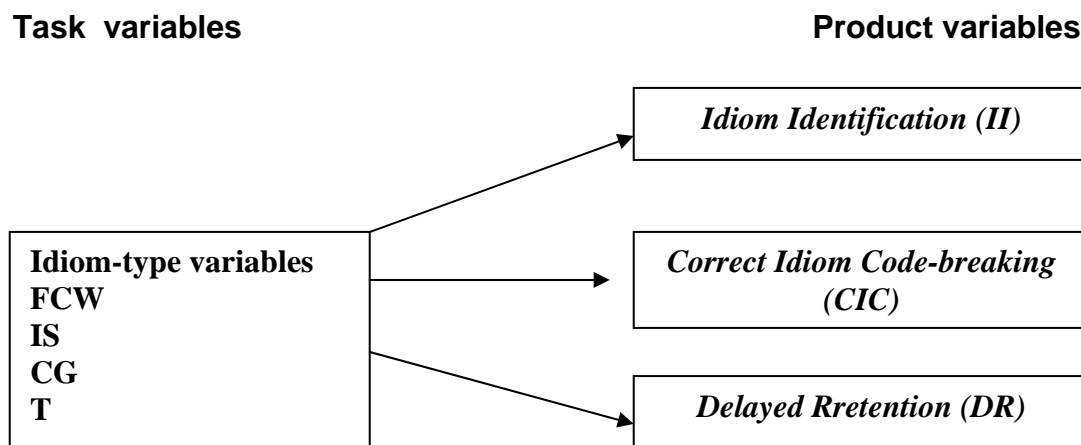
³ Bogaards (2001) focused as well on the familiarity with the component words in the idiom

Universidad de Quintana Roo – Departamento de Lengua y Educación

<http://www.fonael.org> - fonael@fonael.org

type variables included are: a) contextual guessability, b) subjects' familiarity with the words contained in the idioms, c) interlingual similarity, and d) transparency. A series of correlations will allow us determine the variables that could have influenced the results for correct code-breaking and correct retention. See figure 1.1 for the proposed model of interrelationships.

Figure 1.1 Task variables and their effect on product variables



From the design above and the previous literature review, the following hypotheses were posed:

Hypothesis 1: There will be a negative relationship between interlingual similarity and the successful identification of idioms. *Liontas (2001) and Liontas (2002a)*

Hypothesis 2: There will be a positive relationship between interlingual similarity and the successful code-breaking of idioms. *Irujo (1986), Liontas (2002b) and Charteris-Black (2002)*

Hypothesis 3: *There will be a positive relationship between 'interlingual similarity' and the 'familiarity with the component words in the idiom' with delayed retention. (IS) Boers 2000, Boers & Demecheleer 2001, Charteris-Black 2002 & Bogaards (2001). (FCW) Bogaards (2001)*



4. Method

4.1 The subjects

32 Mexican EFL learners participated in the study. They had an intermediate level of English and were in the 6th semester of BA in Applied Linguistics. Their average age was 21 years old and the predominant gender was female. The students have studied English for about 9 years on average. The researcher administered the Nations' vocabulary levels test as a measure of proficiency and the average was 75.5%.

4.2 The materials

4.2.1 The reading text

The text was obtained from a web site called Voice of America⁴. The site provides factual and fiction stories that have been simplified in order to suit different levels of vocabulary. In this case, the text selected falls in the category of texts within the 2,000 most frequent words. The analysis of the text in terms of readability was done through , 'the vocabulary profile' in the website called 'The Compleat Lexical Tutor'.⁵ The vocabulary profile of the text is described in table 1.

⁴ www.voanews.com/SpecialEnglish/adv_search.cfm

⁵ Site: [http:// 132.208.224.131/ The Compleat Lexical Tutor](http://132.208.224.131/TheCompleatLexicalTutor)



Table 1 Vocabulary profile, main study

	No of words	Percentage
First 500 words	824	71.10%
K1 words (1 to 1000)	936	80.76%
Function	593	51.16%
Content	340	29.34%
K2 words (1001 to 2000)	101	8.71%
AWL words (academic)	23	1.84%
Off-list words	99	8.54%
Words in text (tokens)	1159	
Different words (types)	520	
Type-token ratio:	0.45	
Tokens per type:	2.23	
Onlist Tokens	1060	
Onlist Types	452	

4.2.2 The immediate retrospective and elicitation questionnaire (IREQ).

This questionnaire provided information not only about the strategies selected but also the sequence of use of those strategies. It tested the meaning comprehended of the idioms and checked what idioms were previously known by the students. It also provided information about the words in the idioms that were familiar to the students. The code-breaking strategies elicited by the instrument were determined by other studies which attempt to evaluate the lexical processing strategies used by the learners while reading e.g. Alseweed (2000), Fraser (1999). An example of the elicited information for each target idiom is included below.



Be at someone's beck and call

1.- I knew the idiom before reading the text, it means:

2.- I did the following when I read the idiom (1-4)

() Looked it up in the dictionary: *bilingual / monolingual / electronic/ pocket*

() Asked for its meaning to: *teacher / peers/ overheard*

() Ignored it

() Inferred its meaning through: *context /word meaning/ Spanish similarity /metaphor*

3.- The meaning of the idiom is:

4.- I know the meaning of the following words in the idiom:

5.2.3 The interviews

The IREQ described above and the interviews comprise the main instruments that allowed us get data in the form of metalinguistic verbalization the students had at the moment of retrieving specific information about the solving of the task or the solving of specific vocabulary difficulties. In the interviews the text and the questionnaires were presented to the subjects to help them remember more specific information about the strategies used in the identification and code-breaking of idioms. This was enough contextual information to avoid generalizations. These interviews were recorded and transcribed. The students' comments illustrate some of the instances of the results.

The instruments described above were used in the first session where the identification and comprehension data were collected. The following instruments were administered three weeks after in order to collect the data on retention.

4.2.4 The delayed retention test

The delayed retention test has as a main objective the testing of the delayed retention of the idioms that were unknown to the students in the reading. It should be noted that by the end of the first set of tasks all the learners knew which items were idioms in the text but they had never been told the correct meaning or whether whatever meaning they code-broke was correct. The delayed retention test was administered in Spanish and Its questions were expressed as follows:



Be at someone's beck and call

- a) This idiom was included in the text yes () no ()
 - b) This idiom was included in the text but I don't remember its meaning
 - c) This idiom was in the text, it means:
-

The students had to provide an answer to say whether the idiom was or not included in the text (a) or if the idiom was included but they did not remember its meaning (b) or to state that the idiom was in the text and that they remembered its meaning (c). The three questions were asked for every idiom in the text.

5. The measurement of the idiom-type variables

In order to analyze the product data obtained it was necessary to measure the task variables involved in the study. These task variables are specifically the quantification of the idioms in terms of the familiarity of the idiom constituents that the students had at the moment of reading (FCW), the contextual guessability (CG), the interlingual similarity (IS), and the idioms' transparency (T). The procedures followed to carry out this measurement are described below.

5.1 The familiarity of the component words of the idiom (FCW)

FCW is basically the knowledge of the meaning of the individual words of the idiom. This familiarity was determined by the answers provided in the IREQ. The subjects wrote the meaning of the content words contained in each of the target idioms. The score for every idiom was calculated by counting the number of times each of the component words in an idiom were given correctly by all the learners and then dividing the total by the number of content words in the idiom and then by the number of subjects and X 100 to get a percent. For example, idiom (5) was the least familiar of the twenty-one target idioms: '*be at someone's beck and call*', *beck* and *call* were known by 5.11% of the learners, contrary to this, idiom (9), '*be dead on your feet*', was the idiom with the highest familiarity since 90.62% of the learners knew the content words *dead* and *feet*.



5.2 The interlingual similarity of the idioms (IS)

IS was categorized into three types of idioms as suggested by Irujo (1986), Liontas (2002a,b) and Charteris-Black (2002): a) identical idioms, the idioms that were semantically and syntactically equivalent in L2 and L1; b) similar idioms whose meaning and structure was similar in L2 and L1; c) different idioms, idioms that did not have an equivalent or similar form and meaning in L1. The idioms were categorized according to their IS by four native speakers of Spanish of high English proficiency. The four judges were given a form where the target items appeared in a list and the judges had to classify them by giving them a score, 3 for identical idioms, 2 for similar idioms and 1 for different idioms. The translationally equivalent idioms were provided in Spanish: some idioms had two possible equivalences or translations. Hence, the judges scored the two idioms in Spanish and the most frequent was chosen to be codified. The scores were added up and calculated in percent. The following formula was used to get the percent scores: total ratings of all judges X 100 / 12.

5.3 The contextual guessability of idioms (CG)

The CG of the idioms was determined by giving the text to eight native speakers of English. The text was modified in such a way that the twenty-one target idioms were replaced by blanks. The blanks had to be filled in by the native speakers of English. They were asked to write the most plausible idiom, word or phrase they could think of that fitted the context. They were told as well that they could read forwards or backwards to be aware of the context and then fill in the blank. All the guessability evaluators were British, six of them were linguists and two had a different professional profile. A native speaker and the researcher evaluated the responses and assigned two points to the phrase, word or idiom that conveyed almost the same meaning as the original idiom, one point was given to the answers that worked as approximations and 0 points to words or phrases that were totally unrelated to the original meaning of the idiom. The formula to get the



idiom scores of contextual guessability is the following: total of the 8 scores X 100 / 16.

5.4 The transparency of the idioms (T)

The transparency of the idioms was determined by giving the list of twenty-one idioms to four native speakers of English who were all linguists. They were told the phrases were idioms and were asked to indicate the extent to which the meaning of the idioms could be inferred through their constituent words and they were given an example. The three judges categorized the idioms as: transparent, semi-transparent and opaque. Transparent idioms received a score of three points, semi-transparent idioms received 2 points and opaque idioms received one point. The scores were calculated using this formula: total score of 4 judges X 100 / 12.

6. Results

Hypothesis 1:

The hypothesis was not confirmed since no significant correlation was found between II and IS, ($r = -.160$, $n = 21$, $p = 0.488$). The result tends to be negative, it is however non significant. Notwithstanding, we find instances of idioms that were identical and/or very similar and which passed unperceived to the subjects due to the close similarity. The students did not think they were idioms first, because despite their L1 similarity they were unknown to the students and second, the linguistic structure was not noticeable.

S29 After reading we had to identify the idioms by underlining them and the rest was kind of difficult. It was difficult to me because some idioms were similar in Spanish and that made me think, no, this is not an idiom.



S13 I identified thirteen idioms and we already know that there were twenty-one,⁶ so this means that many idioms were not perceived, and even with those that I was able to identify, maybe they have problems with the right meaning. There were some idioms that were so similar to Spanish that I did not take them as idioms. "On the fringes", for example, I didn't think it was an idiom and I treated it in a literal way.

In contrast, we have also found instances where the interlingual knowledge sources (L1-L2 interlingual similarity) of idioms helped to identify the idiom. These cases explain why the correlation was not significant.

'S11 There were some idioms that I was able to identify because they are very similar in Spanish and because in Spanish they are non-literal expressions that imply a different sense'

Three previous studies focused on idiom identification, Irujo (1986) and Lontas (2001, 2002a). The results of idiom identification success obtained in our study can be compared only with Lontas (2001), (2002a) because the subjects were L2 learners and not ESL learners as in Irujo's and also because an explicit identification task was carried out. The recognition of idioms in Irujo (1986) was rather testing the meaning of the idiom. Our results only weakly support Lontas (2002a). In this study also the dissimilar or different idioms in L1 and L2 were detected more often than the other two types. Our study showed the same trend even though no significant difference was found. However, Lontas (2002a) does not present any inferential statistics to show if the difference between different, similar and identical idioms was significant or not. So our result may be the same as his or indeed better since he did not correlate the effects of CG or T on II. The percents of successful identification obtained in the previous studies and in our study can be seen in table 2.

⁶ Learners were not told about the exact number of idioms, but they found it out when solving 'the immediate retrospection and elicitation test', since the idioms were listed in that test, see appendix 3.F.

Table 2 Success in idiom identification (II), comparison with other studies

Study	Different idioms	Similar idioms	Identical idioms	Total success
Liontas (2001)				70.09%
Liontas (2002a) ⁷	96.43%	83.92%	89.64%	89.99%
Our study	53.08%	42.11%	40.54%	42.39%

Hypothesis 2:

The correlation between IS and the correct code-breaking of the idioms was nearly significant ($r = .421$, $n = 21$, $p = .057$ two-tailed). This means that the more similar the idioms were in L1 and L2 the more possibility they had to be correctly code broken. So it suggests that the strategies that used IS were successfully used. Despite these results the hypothesis was not quite confirmed, we include comments made by the learners where the use of IS is displayed.

S24 *'on the tip of the tongue'... and this looks very familiar to Spanish and we use it very often and we say 'I have it here, on the tip of the tongue' with something that we cannot remember at the moment.*

S31 *'cost an arm and a leg' means 'costar un ojo de la cara' (cost an eye on the face). It means it is extremely expensive, it is more or less like in Spanish.*

It is worth checking the possibility of other positive relationships with other idiom-type variables. The correlations show that FCW correlated most positively with the CIC ($r = .438$, $n = 21$, $p = .047$, two-tailed). Neither CG, nor T had any relationship with CIC.

S1, *To a certain extent the meaning of the individual words help as well, "have the nerves" to do something, in a way make us think of courage to do things. I knew the meaning of all the words in the idiom.*

⁷ The percents in Liontas (2002a) in this table include only the Spanish idioms, French and German idioms were not considered.



Hypothesis 3:

An MR test was carried out using the stepwise method with the four idiom-type variables as explanatory variables and DR as dependent variable. Only FCW appeared with a borderline significance: beta: .406, $p = 0.058$. These results indicate that among the various types of idioms concerning CG, T, FCW and IS, the type of idioms that seem to have an impact on the results in correct delayed retention are the idioms categorized by the familiarity that the learners had of the constituent parts of the idioms, (FCW). These results show that the idioms whose content words were more frequently known by the subjects were the idioms that were retained better after a three-week period. The positive correlation between FCW as an idiom-type variable is consistent with the results on the code-breaking strategies that resulted as the best predictors of DR. Inferencing from the meaning of words and also Inferencing from metaphor were predictors of DR and both presuppose knowing the individual words too.

The results discussed above indicate that the hypothesis is partially confirmed since FCW was a predictor variable of correct scores in both CIC and DR in the multiple regression, however IS was not, only FCW seemed to have an important effect. Positive correlations between FWC and the scores of CIC and DR confirm this hypothesis as well. ($r = .438$, $n = 21$, $p = .047$, two-tailed), ($r = .478$, $n = 21$, $p = 0.29$, two-tailed).

7. Conclusions

Interlingual Similarity (IS) was the idiom-type variable that appeared as a predictor of idiom identification, idiom codebreaking and idiom retention in previous studies where the variable was tested in isolation and not together with other idiom-type variables. Our results have shown that **The familiarity with the Component Words in the Idiom** (FCW), appeared as a better predictor of the product measures than IS. The validity of the results is sustained on the multiple correlations of the four idiom-type variables with the product measures.



The results of our study showed that the successful code-breaking of idioms and its successful retention were related to idioms whose component words were previously known by the learners. Hence, a selection of idioms made up of high-frequency component words can be made in order to include those idioms in the lexical syllabus of beginner and lower intermediate learners provided they are important and frequently used by native speakers.

Incorporating a good amount of idioms in very early stages of language learning implies as well raising awareness about the presence of metaphor. It can also be more effective if the idioms presented are those whose component words are already known and those whose interlingual similarity is identical or similar. Our results showed that both types can help learners recognize the metaphor easier.

References

- Alexander, R. J., (1984). Fixed expressions in English: reference books and the teacher. *ELT journal* 38/2 127-134.
- Bahns, J. (1993). Lexical collocations: a contrastive view. *ELT journal* 47/1, 56-63.
- Boers, F. (2000a). Metaphor, awareness and vocabulary. *Applied Linguistics*. 21/4, 553-571.
- Boers, F. (200b). Enhancing metaphoric awareness in specialised reading. *English for Specific Purposes*. 19, 137-147.
- Boers, F. & M. Demecheleer (2001). Measuring the impact of cross-cultural differences on learner's comprehension of imageable idioms. *ELT journal* 55/3, 255-262.
- Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition* 23/ 3, 321-343
- Charteris-Black, J (2002). Second language proficiency: a comparative study of Malay and English. *Applied Linguistics* 23/1, 104-133.
- Cooper, T. C. (1999). Processing of idioms by L2 learners of English. *TESOL-Quarterly* 33/2, 33-262.
- Cornell, A. (1999). Idioms: an approach to identifying major pitfalls for learners. *IRAL* 37/1, 1-21
- Howarth, P., (1998). Phraseology and second language proficiency. *Applied linguistics* 19/ 1, 24-44.
- Irujo, S. (1986a). Don't put your leg in your mouth: transfer in the acquisition of idioms in a second language. *TESOL Quarterly* 20/2 , 287-304.



- Irujo, S. (1986). A piece of cake: Learning and teaching idioms. *English language Teaching*, 40, 236-242.
- Lenon, P. (1998). Approaches to the teaching of idiomatic language. *IRAL* 36/1, 12-30.
- Liontas, J. I. (2001). That's all Greek to me! The comprehension and interpretation of modern Greek phrasal idioms. *The Reading Matrix*, 1/1, 1-31.
- Liontas, J. I. (2002a) Reading between the lines: detecting, decoding, and understanding idioms in second languages. In Sullivan, J. H., ed. *Literacy and the second language learner*. Greenwich, Connecticut: Information Age Publishing.
- Liontas, J. I. (2002b) Context and idiom understanding in second languages. *EUROSLA yearbook* 2, 155-185.
- Lewis, M. (1993). *The Lexical Approach*. London: Thomson-Heinle
- Nattinger, J.R. and J.S. DeCarrico (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Yorio, C.A. (1980). Conventionalized language forms and the development of communicative competence. *TESOL -Quarterly* 14, 433-442.
- Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency in Hyltenstam, Kenneth & L.K. Obler (eds) *Bilingualism Across the Lifespan*. 55-72. Cambridge: Cambridge University Press.

Biodata

Full-time teacher at the Universidad Autónoma de Tlaxcala. BA in ELT (Thames Valley University). MA in Applied Linguistics (UDLA-P). PhD candidate by the University of Essex, UK. Research interests: vocabulary acquisition, learners strategies, all types of action-research projects.
Contact: rosalinangel@hotmail.com; smooth642001@yahoo.com.mx