

Chapter IV. Word analysis

1. Theoretical frame and criteria for the construction of the tool

The field of research for DLA is discourse, more specifically, three levels of it: the narration, speech acts and the word. A tool is available for each level. For the narration, one grid; another two grids for speech acts and a computerized dictionary for the words. The grids allow us to research scenes: 1) those described in the narration (using the corresponding grids), 2) those displayed as speech acts. The scene detected on the speech act level (using the corresponding grids) can be categorized in the broadest terms, taking into account the grid of the narration. Also, the grid of the narration includes two final results: euphoric and dysphoric. The dictionary is particularly useful to detect words, not scenes. But since words and scenes usually coincide, the dictionary can become a useful tool for studying scenes indirectly.

The criterion for collecting words in a specific file (representing a language of eroticism) requires some explanation. the DLA, like other dictionaries, is constructed in a combination of ways: 1) deriving words from concepts (i.e., for GPH, words linked with beauty and for a1, words that express vengeance, revenge, conspiracy, offense, etc.); 2) studying texts in which a certain scene is clearly prevalent (for example, the situation of routine, as a manifestation of UPH, uses words such as “used to”, “almost”, “prudence”, “ambition”, “dignity”, “friendship”, etc.); 3) consulting with judges and advisers, 4) consulting (critically) a thesaurus. The DLA dictionary was constructed with all these means.

Usually dictionaries assign just one semantic value to each word; that is, the word is an expression of one concept. The criticism of linguistic researchers considers that this solution risks mutilating the multiplicity of meanings of each word. The DLA dictionary tries to respond to this criticism by allowing the researcher to consider each word as an expression of more than one concept (eroticism). For each word three concepts are the maximum of options accepted. This decision (more than one meaning for each word) demands a sophistication of the design of the DLA dictionary. Almost all dictionaries function automatically. The DLA also includes an interactive possibility. The interactive use of the dictionary allows the researcher to select one, two or even three options proposed by the dictionary concerning the value of a word in a specific text, taking into account the context of this word (that is, the phrase and/or the narration).

2. Present state of the dictionary

The dictionary has seven files, one for each language of eroticism. In each file, the units are: 1) fragments of words, 2) words, 3) groups of words (for example, compound verbs). The files include a total of approximately 620,000 words, pertaining to approximately 5,000 roots.

Each file has a different number of words, and consequently greater or lesser sensitivity for detecting the corresponding language in the discourse. A calibration system was proposed to balance out these differences:

IL	1.58
O1	2.82
O2	2.10
A1	1.50
A2	1.00

UPH	1.55
GPH	1.07

When a text is analyzed, the program has at least nine functions: 1) it distributes the terms detected into columns corresponding to each language of eroticism; 2) it describes the grammatical features of the words detected; 3) it indicates which words have been detected and which not; 4) it offers the user the different options of erogenous interpretation that the dictionary proposes for a given word and asks the user about the choice: several, all, or none, 5) it reports the number of each term in the whole text, those to which it is sensitive and those which are indicated in each of the columns; 6) it proposes a numerical value for each term detected, corresponding to a calibration index; 7) it presents an overview of erogenous signification (the program has a different color for each language of eroticism) in a given text; 8) it eliminates certain opinions expressed in each column, which refer to those terms that frequently require critical examination; 9) it shows a “pie table” with the percentage of the word for each language of eroticism.

Functions 1,2,3,4,6 and 7 are important in more interactive and handcrafted studies, while a different combination of functions (1,3,5,6,7,8 and 9) is useful when automatic analysis is needed.

Some studies were run in order to detect whether all the languages of eroticism were prevalent in a given text or whether some of them were not assigned their full value by the program. It was verified that all of them have integrated, in some study, the main positions regarding statistical prevalence.

Some of the functions of the program can be clearly exemplified

Example* (partial) of function 1:

* These examples are taken from The Buenos Aires Herald, an Argentine, English-language newspaper whose editorial is also published in Spanish. This example focuses, precisely, on the Spanish version of this editorial. The corresponding English version is:

“Today is the second anniversary of the collapse of Fernando de la Rúa’s Alliance administration with only one question on most people’s minds — will the picket protests called for today result in any replay of the appalling violence of that second-last Thursday in 2001? Yet today’s anniversary is being marked without anybody still knowing what really happened that day (or showing too much curiosity). Yet if we do not know exactly who was responsible, we do know who was not. Two years ago December 19 was spontaneous and peaceful with middle-class anger against confiscatory monetary policies erupting into the famous cacerolazo saucepan-bashing demonstrations — by way of contrast, December 20 was a day of contrived violence with today’s official story of police brutality unable to dispel entirely the suspicions of Peronist goonery to hasten a return to power. But neither the cacerolazos of one day nor the Plaza de Mayo mayhem of the next have anything to do with the pickets — why on earth have they been allowed to dominate today’s anniversary?

“Perhaps because it eminently suits the government that they should be the issue — rather than having to answer awkward questions about the Peronist role in De la Rúa’s downfall, President Néstor Kirchner must surely prefer to pit his personal approval ratings of up to 88 percent against a picket movement whose negative ratings are about as high. Since Peronist goons cannot reasonably be expected to insist on their importance in De la Rúa’s ouster, the only people who could set the record straight here would be the middle-class citizenry manning the cacerolazos. But since the pickets have become an obsessive issue for the middle class in particular, yesteryear’s neighborhood militants are as vulnerable as anybody to amnesia so that the government’s diversionary tactics might well prove successful in terms of diverting attention from the murkier questions about the original December 20.

“But it is a risky strategy whose outcome will not be known until the end of today — playing with the pickets is playing with fire. No point in speculating any further about today’s events but in conclusion, just one question, to those who marched in what was to prove the climax of the “Out with them all!”

IL: 3	O1: 11	O2: 2	A1: 4	A2: 21	UPH: 4	GPH: 10
Value: 4.8	Value: 30.8	Value: 4.2	Value: 6	Value: 21	Value: 6.4	Value: 12
for	question	For	governmen t	Second	Today	One
reproduction	Mind	will give	Protests	Government	Question	Mind
Violence	reproduction		For	Question	Today	For
	Nobody		Violence	Mind	When	will give
	Know			Result		Terrifying
	Occurred			Reproduction		What
	Curiosity			However		Really
	No			Know		Show
	Know			What		Too much
	We know			Occurred		Exactly
	no			Really		
				Or		
				Show		
				Curiosity		
				However		
				No		
				We do not		
				know		
				Exactly		
				responsible		
				We know		
				No		

The example of function 1 shows a clarification of the number of words detected for each language of eroticism (i.e., IL, 3, A1, 4, etc.) and its value (i.e., IL: 4.8, A1, 6). this value (function 6) corresponds only to this specific analysis, since it derives from the application of the calibration index to the number of words detected in each case.

Example (partial) of function 2:

Word	Type of word	Language
Today	Adverb	Phallic urethral
Is	Not found	Not found
The	Not found	Not found
Second	Adjective	Secondary Anal
Anniversary	Not found	Not found
Of	Not found	Not found
The	Not found	Not found
Fall	Not found	Not found
Of the	Not found	Not found
Government	Verb	Primary Anal
Government	Verb	Secondary Anal
Of	Not found	Not found
Fernando	Not found	Not found
De	Not found	Not found
La	Not found	Not found
Rúa	Not found	Not found

wave of disgust with politicians: How do they feel about the anniversary being taken over by groups whose methods are the acme of political manipulation?"

And

Not found

Not found

Example of function 5:

Quantity of words: 447

Words detected : 167 - 37 %

Quantity of opinions: 222 - 49 %

This example shows the quantity of words in the text in analysis, the quantity of words detected in it by the dictionary and the quantity of opinions given by the dictionary. The difference IN the percentage of words detected and of opinions given (in this case, 12%) shows to what extent the dictionary provides more than one opinion for certain words.

Example (partial) of function 8:

IL: 2	O1: 11	O2: 1	A1: 3	A2: 21	UPH: 4	GPH 8
Value: 3.2	Value: 30.8	Value: 2.1	Value: 4.5	Value: 21	Value: 6.4	Value: 9.6
Reproduction	Question	will give	Governme nt	Second	Today	Mind
Violence	Mind		Protests	Government	Question	will give
	Reproduction		Violence	Question	Today	Terrifying
	Nobody			Mind	When	What
	Know			Result		Really
	Occurred			Reproduction		Show
	Curiosity			However		Too much
	No			Know		Exactly
	We may know			What		
	we know			Occurred		
	no			Really		
				Or		
				Show		
				Curiosity		
				sin embargo		
				no		
				sepamos		
				exactamente		
				responsable		
				sabemos		
				no		

Compare these results with the example of function 1.

3. Usefulness, strategies to use the tool, presentation of the results, statistical criteria, sensitivity of the program

The use of the DLA dictionary can: 1) advance results using other tools, 2) enable us to criticize the results of other DLA tools; 3) provide an overview of very extensive material.

The best option for analysis with the dictionary consists in mixing the two strategies: the interactive and the automatic. The complete text can be analyzed automatically, and some selected fragments (the beginning, the last part and certain specific intermediate sections considered relevant by the researcher), interactively. The combination of these two strategies allows the researcher to compare the results and also provides an opportunity to add some new words (detected as expressing a specific language during interactive analysis) to a file of the dictionary.

The results of analysis with dictionaries are usually presented as lines of flux, differentiated for each concept. DLA dictionary results can also be presented this way but, because of its connection with the other two tools (for phrase analysis and narrations), it can be shown as percentages that indicate the relevance of some group of languages and any changes found in other fragments. This option is possible because the unit of research analysis is not the word itself but the phrase or the narration, and the DLA dictionary analyzes the same text as the tools for the study of the phrase or the narration. If not, we would not be able to compare the results of the different studies.

From the statistical point of view some criteria are recommended:

a) which languages of eroticism can be considered to reflect the main features of a certain discourse? The answer to this question depends on the strategy of the researcher's analysis. He could be interested in what is more silent, the state and modifications of the almost mute languages. In this case, he should focus on the results totalling less than 10% of the whole. If the researcher is interested in the major languages present in the text, his decision should be the opposite. In this case, the groups of the 3 or 4 languages that occupy the leading positions and total 70% of the whole is suggested as the best way to depict the dominant features in the discourse. The researcher can include the language that has at least 1) 10% of the whole and 2) 50% of the percentage of the one that was in the last place in the main group. For example, if IL has 11% of the whole and the last position in the main group is held by O2 with 20%, the researcher can also include the former in his list. However, we need to point out that, in some research, the best strategy consists in focusing on the lack of relevance of some languages or even better, of a large group of them, as an expression of the impoverishment of symbolic resources.

b) when can the researcher decide that the difference between two languages is enough to affirm that one occupies the main position and the other, the second? If the results concerning the two languages differ by 3% or more, the researcher can decide to arrange the two languages in a range with differentiated positions. If not, concerning the results of the analysis of the two languages, the researcher should declare a technical draw. For example, if a2 has 30% and GPH 26.40, the first held the main position, but if GPH has 27.40%, the two languages are in a technical draw.

c) comparing the results of the analysis of two texts, when can the researcher decide that the difference between the results on the same language in both *corpuses* is significant or irrelevant? The comparison should take into account the main languages only. When the problem is to detect the percentage of similarities between two languages, the number of words compared has to be at least 50 in the results of both corresponding languages. For example, if UPH has 52 words in one text and in the other only 47, the comparison is impossible. When the problem is to detect the difference, no requirement concerning the number of words detected is necessary. The minimum sufficient for deciding the relevance of the difference is 10% of the calibrated results of the corresponding language. Below this percentage, we can decide that there are similarities between the two texts. If the analysis of text i shows that the main position is for a2, with 30%, and the analysis of text ii coincides and gives this language the same position, but proposes 28% for it, the difference between the two results is irrelevant, since it is under 3%, that is, 10% of 30.

The sensitivity of the program ranges from 30% to 38% of the words in a text. That is, the program detects approximately one out of three words of the text. But every language has some words that cannot be classified because of their high level

of generality. Among these, all the words meaning “the” and a few more. Some research was done for the purpose of determining the percentage of those unclassifiable (from the DLA perspective of the eroticism) words. Studying journalist texts of the daily newspapers *Clarín*, *La Nación*, *La Vanguardia* and *El Mundo* to detect the percentage of those words, the research concluded that between 16.5 and 19% of the words belong to this group. Considering these findings, the sensitivity of the program we see its real value: more than 40% of the relevant words.

4. Systematic use of contextual investigation on the meaning of words: the problem of O1

The researcher can use the contextual analysis of words to test the results of automatic research. Some words require special attention because of their multiple meanings. For example, “como” (“I eat”) is a conjugated verb (“I eat”), a comparative term (“like”), an explicative one (“since”), etc. Functioning as a verb, “como” (“like”) corresponds to IL and O2; used as a preposition, “como” corresponds particularly to A2 and GPH. The dictionary has a function (the 2nd) that informs if the word detected and recognized as a member of one concept is (for the DLA program) a verb, a preposition, etc. If the researcher reads that “como” is used as a verb, he only has to decide between two options (IL and O2), etc.

We have a similar case with “una”, which is a conjugated form of the verb “unir” (“to unite or put together”), but is most often used as an article (“a”). As an article it does not belong to any concept, but the dictionary detects it as a verb in GPH. Nevertheless, usually the opinion of the dictionary is erroneous, because the term in the text is most frequently an article rather than a verb.

When the researcher wants to advance quickly, avoiding the interactive strategy, he can use a function (the 8th) of the program that allows him to eliminate all these more complex opinions. If not, he has to decide what to do with each of these problem words. Nevertheless, the difference between the two results (interactive and automatic) is more or less 1%. Also, interactive analysis allows the researcher to detect some compound words that the program cannot spot (for example, “golpearlo” (“to beat him”) = “golpear” + “lo”; “golpear” is usually an expression of a1, and “golpearlo” is too). These additions and rectifications change the results of the analysis by less than 1%.

But the greatest value of the analysis of words-in-context with the DLA dictionary lies in a different field, linked with the decision concerning the results of o1 and a2 languages. Usually each word has multiple meanings that are restricted by the context. Frequently all the meanings that the DLA dictionary proposes for each word should be accepted. But in certain situations the researcher needs to choose, particularly concerning O1 language. This language includes terms connected to thinking and other intellectual activities. The same happens with A2 language. The difference is that O1 refers to abstract, perhaps mystical thinking that may even reject reality testing, while A2 refers to concrete thinking and traditional knowledge. Usually the person who prefers o1 tries to reject A2 as an opponent to be convinced, even destroyed. The files of the program of o1 have a great quantity of words in common with the file of A2. Therefore, when studying the text the researcher must decide whether the two languages are co-present. Because of the high value given by the calibration process, O1 frequently takes the pole position from the statistical point of view. But this evaluation can be incorrect: perhaps O1 is not expressed on the other two levels, neither in the narration nor in the speech acts, and all the words detected

as members of O1 coincide with words detected as belonging to A2. If O1 does not have specific words (not coinciding with A2), such as telekinesis, astronaut, miracle, revelation, mysticism, genius, etc., and if on the levels of scenes (speech acts and narration), equivalents (e.g., the scene of mystic revelation, or the scene of the emergence of an extraordinary idea in the mind of a genius) do not appear, the opinions concerning O1 on the level of words has to be rejected. In this case, the decision depends on two combined facts: 1) the absence of specific words representing O1, differentiated from A2, 2) the absence of speech and narrations expressing O1.

5. Criticisms and limitations

Computerized dictionaries have received general criticism:

1) the tool ignores the relevance of the semantic value of the words because of their insertion in a specific context (phrase, etc.) the DLA dictionary tries to deal with this criticism by including more than one semantic value for the words and by offering the researcher a complementary interactive function that allows him to view the context of the phrase for each word investigated.

2) the tool cannot analyze metaphors and other semantic rhetorical resources. But the phrase “the pearls in your mouth” (which contains an old metaphor) can be analyzed in its textual value: “pearl” is detected as a jewel (GPH), and “mouth” is detected too. What the dictionary cannot perceive is that the word “pearl” has metaphoric value. But this kind of analysis belongs to the phrase level, not the word level. And on the phrase level, the corresponding grid has some items to guide the researcher.

3) the tool cannot detect slang uses of words and some historical changes in the meaning of the terms. But the terms don't change so fast (if not, we could neither write this phrase nor our reader understand it). Actually, the same criticism is valid for all types of dictionaries, while the slang and regional meanings of certain words can be investigated using the interactive function of the DLA dictionary. In all idioms cocaine has vernacular names, that the dictionary cannot detect; but if the researcher asks the dictionary for the semantic value of cocaine he can find it (mostly IL). He needs only to add to the slang word the same semantic value that the dictionary proposes for the “official” word, like cocaine.

Also, the results found with the DLA dictionary allow the researcher to detect what languages prevail, though not which has logical dominance. Sentence I: “I prefer to dress up nicely and receive gifts, but I have to clean the library” contains the same words as sentence II: “I have to clean the library, but I prefer to dress up nicely and receive gifts”. The dictionary detects that “I prefer to dress up nicely and receive gifts” has a great dominance of GPH, and that in “but I have to clean the library” A2 prevails. Perhaps statically speaking GPH is dominant on the level of words, but on the level of the phrases the analysis has nuances: in sentence I, A2 prevails, in the sentence II, GPH. The statistical prevalence of some result of word analysis needs to be contrasted with the results of the analysis of the narration and of the phrase. Nevertheless, the statistical analysis of the words usually has a strong influence on decisions concerning the dominance of certain languages over the rest. Perhaps the researcher concludes that sentence I is an expression of dysphoric results for GPH and a euphoric result for A2, and that in sentence II the solution is the opposite. But in the whole text the researcher can detect that the most important language is GPH (statistically dominant too on the level of words), and that prevalence includes a

dysphoric version (as in phrase I). So, the results of the analysis of the words have a complex relationship with the results of the other two levels of investigation.

Another limitation concerns the detection of A1. Very usually wishes of vengeance and justice are disguised and appear in phrases having a second intention. In phrase level A1 can be detected as threats, jactancy, triumphalistic mockery, etc., but no word indicates them. Usually A1 is detected more in the terrain of narration or para-verbal components than in the other two (phrase and word levels). In consequence, concerning A1 a difference between the corresponding outcomes of analysis can appear.

6. Applications

6. 1. Journalism sections

Some research into different sections of the most important newspaper of Buenos Aires, the *Clarín*, was carried out using the DLA. The sections studied were: i) recipes, ii) social notes, iii) economy, iv) wine & dining, v) police information, vi) sports, vii) fashion, viii) obituaries, ix) political opinion page, x) international news, xi) the world of computers, xii) national political news. The result of the analysis with the DLA dictionary indicates similarities and differences between them:

I. Recipes	II. Social notes	III. Economy	IV. Wine & dining	V. Police	VI. Sports
1. GPH 34.07	1. GPH 27.68	1. A2 27.39	1. GPH 28.31	1. UPH 25.90	1. GPH 26.48
2. UPH 20.58	2. UPH 25.23	2. IL 24.28	2. A2 27.71	2. A2 23.39	2. UPH 26.10
3. A2 12.89	3. O1 16.48	3. GPH 19.30	3. UPH 23.43	3. GPH 14.81	3. A2 17.91
4. IL 11.38	4. A2 14.82	4. UPH 13.98		4. A2 14.68	4. O2 14.10

VII. Fashion	VIII. obituaries	IX. Political opinion page	X. International news	XI. the world of computers	XII. National political news
1. GPH 34.08	1. O2 37.19	1. A2 28.23	1. A2 33.71	1. O1 32.08	1. A2 33.79
2. A2 24.08	2. IL 16.27	2. UPH 18.57	2. UPH 26.12	2. A2 31.12	2. UPH 19.33
3. UPH 18.06	3. UPH 14.68	3. O2 16.19	3. O2 16.68	3. UPH 10.41	3. GPH 16.12
4. IL 15.95	4. GPH 11.45	4. GPH 15.61			

These results, compared to those obtained by analyzing the same sections of the other major Argentine newspaper, may be interesting. In *La Nación* the results are:

I. Recipes	II. Social notes	III. Economy	IV. Wine & dining	V. Police	VI. Sports
1. GPH 38.89	1. GPH 29.71	1. A2 22.44	1. GPH 31.14	1. A2 31.24	1. UPH 21.33
2. UPH 21.66	2. UPH 19.86	2. O2 20.10	2. A2 22.21	2. O2 21.87	2. O2 18.66
3. IL 13.25	3. A2 17.79	3. IL 19.29	3. UPH 17.21	3. UPH 20.75	3. A2 18.35
4. A2 13.04	4. IL 15.74	4. GPH 17.83			4. IL 15.85

VII. Fashion	VIII. Obituaries	IX. Political opinion page	X. International news	XI. The world of computers	XII. National political news
1. GPH 32.35	1. O2 37.48	1. A2 31.61	1. A2 29.99	1. O1 34.13	1. A2 35.61
2. A2 25.09	2. UPH 19.01	2. UPH 16.98	2. UPH 20.33	2. A2 25.93	2. UPH 19.69
3. UPH 17.38	3. IL 15.86	3. A1 15.22	3. GPH 17.55	3. GPH 12.73	3. A1 13.90
		4. O2 14.77	4. O1 14.76		4. GPH 13.88

In the *Clarín* some differences in the results of analysis of the first two positions are not relevant (sections II, VI, XI). Nevertheless, we notice that with the exception of Police (V) and Sports (VI), the first language detected for a section is the same in both newspapers. I (Recipes) and II (Society) has the same prevalence IN the two main languages, and IV (Wines & dining) and VII (Fashion), too. VIII (obituaries) and XI (the world of computers) are very different from the rest of the sections.

This research indicates that the ten sections where at least the first language in both newspapers coincides can be categorized into three great groups: 1) those in which the first position is occupied by GPH (I, II, IV and VII), 2) those in which A2 is in the main position (II, IX, X and XII), and 3) those with a singular first position (VIII and XI).

In the first group the promise, the offering of nice moments, etc. are prevalent; in the second, objective information is predominant. In the third, the importance of expressing feelings (O2) is especially clear in VIII, and of abstract thinking (O1) in XI.

Another investigation studies the results of analysis with the DLA program applied to the same section (the Political opinion page) in four newspapers published the same day. The newspapers are: *La Nación*, *Clarín*, *Página 12* and *Buenos Aires Herald*. The analysis of these texts shows the same results for all of them in the first two positions: 1. A2, 2. UPH. These results of analysis with the program on the word level coincide with the results of analysis on the levels of the phrase and Especially the narration. The writer shows the same conception of space, group, values, *Weltanschauung* in the narration scene, which can be inferred by the analysis of the words. So, the agreement of these results enable us to affirm that the same sections in different newspapers have certain features in common, from the point of view of the languages, concerning the representation of values, groups, space and *Weltanschauung*.

A complementary investigation of the political opinion page in the same newspapers (*Clarín*, *La Nación*) several months later yielded conclusions that tally very well with the previous findings.

Political opinion page (*Clarín*)

	15/II/04
A2	25.79
UPH	19.90
O2	16.04
GPH	14.21

	22/II/04
A2	29.97
UPH	20.93
GPH	13.62
O2	12.68

	29/II/04
A2	29.92
UPH	20.09
O2	14.29
GPH	13.29

	08/II/04
A2	31.35
UPH	17.17
GPH	15.17
O2	14.65

Political opinion page (*La Nación*)

	15/II/04
A2	28.73
UPH	19.29
GPH	15.43
A1	12.56

	22/II/04
A2	31.97
UPH	20.40
GPH	13.36
O2	13.26

	29/II/04
A2	31.72
UPH	20.31
A1	14.89
GPH	12.98

	07/III/04
A2	33.02
UPH	21.01
GPH	19.34
A1	17.73

Also, the study of some sections (Recipes, Fashion) in a Spanish newspaper (*El Mundo*) shows these results:

Recipes

	14/12/03
GPH	32.96
A2	17.50
UPH	16.81
O2	15.52

	04/01/04
GPH	37.19
UPH	18.46
IL	14.50
A2	11.32

	29/02/04
GPH	29.58
A2	20.22
O2	17.70
IL	15.18

	07/03/04
GPH	36.65
UPH	20.55
O2	19.29
IL	11.61

Fashion

	11/01/04
GPH	42.46
A2	21.13
IL	11.40
O2	10.82

	29/02/04
GPH	41.65
A2	24.33
UPH	12.72
O2	10.85

	07/03/04
GPH	31.25
A2	24.25
UPH	16.11
IL	14.86

	07/03/04
GPH	40.93
UPH	20.11
IL	15.11
A2	12.97

In reference to national political news, the analysis of the Spanish newspapers *La Vanguardia* and *El Mundo* yields these figures:

<i>El Mundo</i>	
1. A2	45.87
2. GPH	15.84
3. UPH	15.55

<i>La Vanguardia</i>	
1. A2	46.13
2. O2	16.91
3. UPH	15.88

These results agree with the findings of the study of the same sections in the Argentine newspapers. Therefore, we can extend the previous conclusions to the same section in other newspapers.

Nevertheless, sometimes certain subjects determined a specific change in the results of analysis. For example, in the middle of 2001, for a month, the author of the political opinion page of *Página 12* focused his attention on economic issues. Therefore, the results of the analysis were different from those of more recent texts.

June-July 2001

1. A2

2. IL

Dec. 2004

1. A2

2. UPH

We can infer that each section of the newspaper has an internal code, derived from an implicit social contract between writers and readers. The latter expect to find some kind of style, and the former display that style, which the internal criticism of the reviewers of the newspaper controls and supervises. Some stylistic variation is allowed, depending on the subjects presented, and perhaps in some other conditions (for example, the supposed social feelings of the readers). Some sections have two strategies of exposition. For example, in the world of computers section, sometimes "objective" information prevails; in consequence A2 has the pole position and O1 takes second.

Another example of these variations can be seen by comparing the results of some Spanish and Argentine newspapers dealing with the same subject. The Spanish (*El Mundo*, *La Vanguardia*) and The Argentine (*Clarín*, *La Nación*) newspapers of 16/iii/04 included information on the terrorist bomb attacks in Madrid and their effects on parliamentary elections in Spain.

	<i>El Mundo</i>
A2	28.07
A1	21.52
GPH	17.57
UPH	16.43

	<i>La Vanguardia</i>
A2	30.55
A1	18.79
UPH	17.72
O2	17.08

	<i>Clarín</i>
A2	26.19
UPH	24.68
A1	15.40
O2	15.10

	<i>La Nación</i>
A2	27.75
A1	18.66
O2	18.08
GPH	15.35

In this case, the relevant position of A1 is relatively unusual in national political news and the international news sections. These variations in the results depend on the subject discussed and perhaps also the supposed social sentiment of the readers in connection with scenes of injustice and abuse of the power.

Also, considering the sensitivity of the DLA dictionary applied to the different sections, we notice that the program detects about 35% in nine of the parts. Only obituaries and sports have lower percentages. The use of certain terms (almost a slang vocabulary) and names in the sports section and the prevalence of names of the honored dead, etc., in the obituaries section explains the lower percentage of DLA sensitivity.

6. 2. Translation

Could research with the DLA concentrate on translated texts? Some research focused on a delusional mystic book (*Neuropathic memories*, written by Schreber). The original German text has two Spanish translations (I and II). The two versions of a chapter of that text ("on hallucinations") was analyzed using the DLA dictionary. The research mixed automatic and interactive strategies: both versions of the entire chapter were analyzed automatically (about 3,850 words), and the start and the end (about 500 words each), interactively. The results of both studies showed very little difference. The research compares the difference between the percentage given by the program for each eroticism after studying both versions. For example, the analysis of version I shows that o1 has 26.67 % of the whole, while in the analysis of version II it has 25.70%. The difference between the two results is 0.97%, as shown in the table below. Sometimes, concerning a certain language of eroticism (i.e., IL) the analysis of one version (I) detected x% more than the analysis of the other (II); but the same analysis arrived at an inverse conclusion (that is; for version I x% less than the analysis of the version II was detected). The tables below show the amount of difference between version I and version II, but not which version has more and which less words detected. The study reveals an alteration in the results: for one language of eroticism there is x % more in version I, and for another, x% more in version II.

In the automatic analysis the two studies agree relatively in the range of the prevalences:

	Version I	Version II	Difference
1. O1	26.67%	25.70%	0.97
2. A2	20.41%	20.93%	0.52
3. GPH	16.05%	17.26%	1.21
4. UPH	15.24%	15.77%	0.53

Between the third and fourth position in Version I there is a technical draw. Interactive analysis of the start of the two version yields these results:

	Version I	Version II	Difference
O1	31.96%	34.45%	2.49
A2	25.11%	25.33%	0.22
GPH	14.79%	15.34%	0.55
UPH	11.32%	12.74%	1.15

The interactive analysis of the last part of the same text arrives at these results:

	Version I	Version II	Difference
O1	36.29%	32.85%	3.44
A2	18.96%	17.48%	1.48
GPH	14.40%	15.73%	1.33
UPH	12.67%	15.09%	2.42

In the interactive analysis, the greatest differences between the two versions correspond to O1 and reached 3.44%, less than the corresponding 10%. In the analysis of the last part of the text, the difference between the two results in UPH arrives at 2,42%, which is more than the corresponding 10%. In both texts the number of words belonging to UPH is under 50. Consequently, this result is irrelevant. Also, the differences between automatic and interactive analysis of the same fragment were 1%. These differences are also not significant; in consequence, from the perspective of the DLA analysis, the two versions are equivalent and can be studied by this method. Obviously, the translated text lost: 1) the phonological values of the original version; 2) some syntactic nuances and 3) the semantic resonance of several words. But, from the perspective of the languages of eroticism, several important features are preserved. hypothetically, one version can provide, as a translation, "pleasant", another one, "agreeable", and even a third one, "nice". The three have different phonologic value, even different semantic resonance, but, from the perspective of the dictionary, all three are detected as expressions of GPH.

Incidentally, concerning the analysis of Schreber's text, the results of the investigation of the narration and the phrase levels showed great agreement with the results of the analysis of the words using the dictionary (see sector B, chapter I).