Universidad de las Américas Puebla

Discourse Across Borders:

A Quantitative Analysis of Mexican and American Media Discourses

"…if thought corrupts language, language can also corrupt thought."
George Orwell, *Politics and the English Language (1946)*

Tesis sometida a la

Facultad del Departamento de Lenguas

Como Requisito Parcial para la Obtención del Grado

Master en Applied Linguistics

por

Jesse Sinclair Owens

(Primavera 2012)

La Tesis ha Sido Leída por los Miembros del Comité de Tesis de

Jesse Sinclair Owens

Y Considerada como Requisito Parcial para Obtener el Grado de

Master en Applied Linguistics

Jurado calificador:

Presidente: Dra. Magdalena Mejía Gómez

Secretaria y directora de tesis: Dra. Brita Banitz

Vocal: Dra. Louise Greathouse Amador

Cholula, Puebla, México a 10 de mayo de 2012

Acknowledgments:

I would like to thank my parents as well as the rest of my family (tanto en los Estados Unidos como en México) for their constant support.

Dedicated to Esther Francisco, without whom none of this would have been possible.

Abstract

Critical Discourse Analysis (CDA) is a well-known form of discourse analysis which is built around the examination of the ideological, social, historical and cultural components of discourse. The field has become well-established since the early 1980s and has—as a result—been the target of criticism from linguists as well as scholars from other fields. Much of this criticism has come about as a result of a tendency among CDA researchers to study 'cherry picked' data (Mautner, 2009) and to draw conclusions which are heavily influenced by personal ideologies (Poole, 2010).

The research presented here provides an alternative to 'traditional' forms of CDA research. Using elements of Methodological Synergy (Baker et al., 2008) and Semantic Prosody (Morley & Partington, 2009; Louw, 2008), a research methodology was designed and used to study English and Spanish newspaper discourse regarding 'drug-related violence' in the United States and Mexico. Through the use of Methodological Synergy and Semantic Prosody, it is hoped that both selective data collection and ideologically tinted analysis can begin to disappear from CDA research, resulting in a more thorough and objective form of linguistic research.

Abstracto

El Análisis Crítico del Discurso (CDA, por sus signos en inglés) es una forma común de investigación lingüística que está enfocada en el análisis de discurso tomando en cuenta factores ideológicos, sociales, históricos y culturales dentro del mismo discurso. Desde sus inicios en los años 80, el campo ha evolucionado y ha sido utilizado con más y más frecuencia con el paso de los años. Como resultado, CDA también ha sido blanco de críticos por parte de lingüistas y también por investigadores de otras áreas. La mayoría de estas críticas han sido en contra de la tendencia de muchos investigadores del CDA de utilizar datos seleccionados sin tomar en cuenta la necesidad de mantener objetividad (Mautner, 2009) y de llegar a conclusiones que están basadas en las ideologías de los mismos investigadores (Poole, 2010).

La investigación presentada aquí tiene como propósito proveer una alternativa a formas 'tradicionales' de realizar investigaciones en CDA. Usando elementos de *Methodological Synergy* (Baker et al., 2008) y *Semantic Prosody* (Morley & Partington, 2009; Louw, 2008), una metodología fue diseñada y después utilizada para estudiar discursos mediáticos en los periódicos de México y de los Estados Unidos que trataron de la 'narco violencia' en ambos países. Por medio del uso de *Methodological Synergy* y *Semantic Prosody* se espera enfrentar los problemas que proceden de formas tradicionales de la recolección de datos (en el CDA) y el análisis basado en ideologías para con ella tener una visión más objetiva y completa de realizar investigaciones del CDA.

Table of Contents

Acknowledgments…….....iii

Abstract……..iv

Table of Contents……v

List of Tables……vii

List of Figures……..viii

1 INTRODUCTION……..1

1.1 Background/Description of Methods Used……...6

1.1.1 Critical Discourse Analysis……6

1.1.1.1 Approach/Characteristics……8

1.1.1.2 Background and Shortcomings……13

1.1.2 Corpus Linguistics…….....15

1.1.3 Semantics……17

1.1.3.1 Semantic Prosody……..19

1.2 New Method……21

1.2.1 Current Study……...23

2 LITERATURE REVIEW……29

2.1 Critical Discourse Analysis…….29

2.2 CDA Studies of Media……32

2.3 Corpora and CDA……...40

2.4 Application of a Semantic Approach……..44

3 METHODOLOGY……...52

3.1 Overview of the Methodology Employed……...53

3.1.1 Pilot Study……58

3.1.1.1 Texts……..62

3.1.1.2 Text Selection Criteria……..64

3.1.2 Sample Selection and Corpus Construction…….66

3.1.3 Statistical Analysis of Corpora…….73

3.1.4 Node and Collocation Selection……...75

3.1.4.1 Sorting of Nodes and Collocations……...79

3.1.4.1.1 Node Words……...79

3.1.4.1.2 Concordance Analysis……82

3.1.5 Final Analysis of Corpora……84

4 RESULTS……..87

4.1 Frequent Node Words…….88

4.2 Frequent Equivalent Nodes……..….101

5 CONCLUSIONS……115

5.1 Frequent Node Analysis……116

5.2 ‘Shared Concept’ Analysis……119

5.3 Discussion…….121

5.3.1 Corpus Issues…….122

5.3.2 Strengths and Weaknesses…….123

REFERENCES……..126

Appendix A…….135

Appendix B…….138

Appendix C……..……139

Appendix D…….140

Appendix E……..141

Appendix F……..142

List of Tables

Table 1. Search terms used for English corpus piloting……59

Table 2. Search terms used for Spanish corpus piloting……...60

Table 3. Criteria for inclusion in or exclusion from English and Spanish corpora……...67

Table 4. Lexical and topical elimination criteria for English and Spanish corpora……..72

Table 5. Articles eliminated from final corpora……...…….73

Table 6. Distribution of lexical items according to frequency in the English corpus……...76

Table 7. Distribution of lexical items according to frequency in the Spanish corpus……...77

Table 8. Top ten most frequently occurring nodes in the English corpus…….80

Table 9. Top ten most frequently occurring nodes in the Spanish corpus……80

Table 10. Top ten most frequent LIs in both corpora……...89

Table 11. Collocates of MEXICO in the English corpus……..92

Table 12. Collocates of MÉXICO in the Spanish corpus…….93

Table 13. Collocates of BORDER in the English corpus…….95

Table 14. Collocates of FRONTERA in the Spanish corpus……95

Table 15. Collocates of PAÍS in the Spanish corpus……98

Table 16. Collocates of COUNTRY in the English corpus……..98

Table 17. Collocates of POLICE in the English corpus……..105

Table 18. Collocates of POLICÍA in the Spanish corpus……...105

Table 19. Collocates of NEW in the English corpus……..109

Table 20. Collocates of NUEVO in the Spanish corpus…….110

Table 21. Collocates of AUTHORITIES in the English corpus…….111

Table 22. Collocates of AUTORIDADES in the Spanish corpus……...111

Table 23. Collocates of MILITARY in the English corpus……112

Table 24. Collocates of EJÉRCITO in the Spanish corpus…….112

List of Figures

Figure 1. Process for inclusion in or exclusion from final corpora……...65

Figure 2. Number of articles per day included in English and Spanish corpora……...70

Figure 3. Sample KWIC presentation from English corpus (MEXICO)……..82

Figure 4. Concordances of *security* with the node word BORDER in the English corpus...97

Figure 5. Concordances of *vive* with the node PAÍS in the Spanish corpus…….99

Figure 6. Concordances of *officers* with the node POLICE in the English corpus…….107

1. Introduction

Critical Discourse Analysis (CDA) is a branch of Discourse Analysis (DA)[1] which focuses on the connections and interactions between language use, ideology, power, discourse and sociocultural change (Fairclough, 1995). As a method of analyzing these issues CDA has existed and been prominently used for long enough to establish itself as a recognized and generally respected branch of Applied Linguistics research. CDA has not only helped to expand the broader linguistic field of DA, but has given rise to a few widely-used DA approaches such as Ruth Wodak's *Discourse Historical Analysis* (DHA) (see Wodak, 2007 for discussion) as well as a variety of CDA approaches which examine issues such as racism and discrimination (see, e.g., van Dijk, 1988) and issues of ideology and power (see, e.g., Fairclough, 1995). Nonetheless, due to CDA's patent connection to social and political issues—both at the level of commencing research and at the level of carrying it out (see Fairclough, 1995; Carvalho, 2008)—CDA research often spawns its own discourse, featuring the same sorts of underlying critical linguistic characteristics that it sets out to analyze (see Poole, 2010). Despite the fact that CDA is presented as a way to bring underlying ideological currents in discourse to light, it often harbors a discourse of its own in the form of its analysis and conclusions. The fact that CDA research is carried out and written by an individual with ideological leanings and that it is approached with specific ideological goals in mind results in a text with its own ideologically marked discourse—often similar in discursive features to the text(s) being analyzed. Because of this, linguistic claims put forth by CDA researchers are often diluted by the social and political commentary present on the part of the researchers themselves (Prentice, 2010).

[1] For the sake of economy within the text, 'CDA' and 'DA' will be used throughout the present text when referring to these fields (*Critical Discourse Analysis* and *Discourse Analysis*, respectively).

The current study is intended as a means through which to propose a new approach to traditional CDA-style research. This approach is intended to allow for a more balanced, more objective, and less ideologically influenced manner of conducting CDA research. The methodology being proposed and described here is effectively a combination of the *Methodological Synergy* approach (Baker et al., 2008) and some elements of *Semantic Prosody* (SP) analysis (see Zhang, 2010; Morley & Partington, 2009; Louw, 2008 for general discussion). This research, whose theoretical and methodological underpinnings are discussed in more detail below, is focused on the analysis of Mexican and American newspaper texts reporting on issues related to the drug trade and resulting violence between the two countries. Specifically, the current investigation looks at print media discourse related to the violence spawned by the drug trade. For the sake of consistency, this topical focus will henceforth be referred to as 'drug-related violence.'

The analysis of 'drug-related violence' in newspaper reporting was approached by embracing the possibility that carrying out CDA research which is not ideologically influenced or overtly subjective is viable through the use of a methodological approach which eliminates many of the ideologically tinged elements of traditional CDA investigations. Through the use of corpus linguistics tools, the statistical analysis of lexical features, and SP analysis, media discourse was compared and contrasted between the United States and Mexico. Through the use of this 'Methodological Fusion,' a representation of each country's print media environment with regard to current 'drug-related violence,' will be analyzed and the methodological approach presented here will be found to be either be a viable approach to future CDA-style studies, or not.

Using the current research project as a starting point, it is proposed that it is possible to critically analyze the use of language in textual discourses from a more objective perspective than has been utilized in past CDA studies. This can be accomplished by employing a positivist approach to traditional CDA research. Through the use of a combined methodology based on corpus linguistic analysis and semantic prosody analysis, it will be possible to focus on underlying discourse characteristics across two corpora. This will additionally permit research to move away from the subjective political and ideological commentary which characterizes much of traditional CDA research and which has made CDA such a hotly debated issue in many academic circles (see, e.g. de Beaugrande, 2001; Stubbs, 2001a; Widdowson, 2001a; Widdowson, 2001b).

Many of the approaches to and conceptualizations of CDA and its place in modern applied linguistics research informed the present study. These approaches to CDA research are criticized in the present work because in many respects they are representative of highly subjective understandings of DA. Despite the fact that a critical reading of much of the past few decades' worth of CDA research served as the motivation for the present investigation, it should be made clear from the beginning that some readings of CDA, such as Fairclough's (1992) 'assessment' of discursive and social change or Bourdieu's (2006) examination of discursive power relationships, in which, "utterances [are seen to] receive their value (and their sense) only in their relation to a market, characterized by a particular law of price formation," (p. 481), are not overtly misguided. Nonetheless, the importance necessarily placed on the author's subjective interpretation in these studies—what Wodak (2000) calls being 'self-reflective'—detracts from the analysis which comes as a result of the research itself. As a response to this preponderance of subjective, ideologically based

analysis within the field, the present study presents a combination of methods which has the final goal of establishing an alternate approach to traditional CDA studies, and which is as near to being wholly objective as possible.

The application of some of the approaches which will be used in this study is not new, *per se*. For example, some authors (most notably van Dijk, 1988) have prominently applied semantic theory to areas of CDA in the past. More recently, there has been an increased interest in the use of corpus linguistics in CDA research. While research utilizing this approach initially lacked methodological cohesion (see Orpin, 2005), the use of corpora in CDA research has become much more consolidated with the introduction of the Methodological Synergy approach proposed by Baker et al. (2008) and subsequently employed by Salama (2011) and Freake, Gentil and Sheyholislami (2011). However, these types of approaches to traditional CDA research are not very common and inevitably employ many of the same subjective discourses in their analyses as do other CDA studies. For the purposes of the present study, this discourse 'circularity' (Stubbs, 1997) is seen as the result of the two basest tools used in carrying out a CDA-based study: the sample selection and the analysis itself. Because ideological factors necessarily inform both of these methodological steps, it becomes difficult to respond to weaknesses in CDA research without addressing both of them, something that has yet to be fully embraced in the field (for a simple discussion of ideology's role in CDA research, see Wodak & Meyer (2009)). While some authors have chosen to more heavily focus on sampling issues (Freake et al., 2011) and others have placed more weight on issues of analysis (Prentice, 2010), few have attempted to address both of these issues in a single methodology. Although a more balanced answer to traditional CDA has been proposed and implemented by some authors

(Salama, 2011; Baker et al., 2008), it has yet to be applied to discourse in the same way as has been traditional CDA. That is, while many studies have focused on placing more or less weight on both of these points, they have either fallen short of giving equal weight to both, or have resulted in impractical methodologies—thus conflicting with one of the principal goals of CDA. These extremes in methodologies can be seen with particular clarity in many of Fairclough's studies, which have focused heavily on sample selection at the cost of a balanced, objective analysis (see Poole, 2010; Fairclough, 1993) and in studies based on corpus assisted discourse studies (CADS) (see Freake et al., 2011; Salama, 2011; Baker et al., 2008) which have sacrificed practicality in the name of balanced analysis.

Even though each of these varied focuses on CDA research has taken great strides to make their findings more objective, they have each only addressed part of the issue raised when examining general subjectivity in CDA research. In van Dijk's (1988) exploration of the application of DA, for instance, van Dijk uses semantic theory in order to improve the meaning analysis which is present in many CDA studies. However, in doing so, the data collection component of the research remains the same. Similarly, in the Methodological Synergy approach (Baker et al., 2008—discussed in more detail below) the researchers address the subjectivity traditionally present in data collection, but ignore the subjective pitfalls present in the data analysis. In response to this seeming methodological vacuum, the current project intends to piece together existing concepts within CDA, corpus linguistics, and semantic theory, with the end goal of establishing a more empirically sound form of CDA than has heretofore been utilized in the field.

1.1. Background/Description of Methods Used

In the interest of effectively establishing the precedents which informed the current research, it is first necessary to discuss and describe the theoretical and methodological foundations which will make this research possible. As has already been noted, the current project was based on a fusion of multiple approaches to DA that have already been used within linguistics. The areas (CDA, Corpus Linguistics and Semantic Prosody) are presented below and their theoretical and methodological backgrounds are briefly discussed.

1.1.1. Critical Discourse Analysis

CDA is a form of DA which came to prominence in the late 1980s and early 1990s (Wodak & Meyer, 2009; Blommaert and Bulcaen, 2000; Fairclough, 1995). The approach has since gained considerable popularity and is championed by authors such as Norman Fairclough, Teun van Dijk, and Ruth Wodak, among others. Generally, the approach analyzes written or spoken discourse by viewing and interpreting it through a lens of political culpability and social justice. As opposed to traditional forms of DA, CDA views discourses as reflections of the society in which they are created. In this way, CDA endeavors to bring to light underlying discourses within greater linguistic trends. This approach, and the goals that inherently come with it, make CDA a sort of tool with which researchers may 'uncover' discourses which affect the public in their everyday lives, what Fairclough (1995, p. 1) calls, "…a resource for people who are struggling against domination and oppression in its linguistic forms." Indeed, the examination of 'domination' presented through discourse is one of the pillars of CDA research. This

'domination' is often seen as being connected to different groups within a society who put forth and participate in discourses. Wodak and Meyer (2009), for instance, view this connection as being predominantly caused by power—in their view an extremely important facet of CDA. In this sense, those in power are seen to be responsible for social inequality, and CDA investigators are seen to be, "…interested in the way discourse (re)produces social domination, that is, the power abuse of one group over others, and how dominated groups may discursively resist such abuse," (p. 9).

Because CDA studies offer findings so closely tied to interpretations of socio-political matters (i.e. 'power,' history, race issues, politics, etc.), they are necessarily characterized by interdisciplinary research and, thus, informed by a wide range of approaches and methodologies. All of this is done while viewing language use and its presence in society through a critical lens. As Wodak points out, "*critical* does not mean detecting only the negative sides of social interaction and processes and painting a black and white picture of societies. Quite to the contrary: *Critical* means distinguishing complexity and denying easy, dichotomous explanations. It means making contradictions transparent," (2000, p. 186, emphasis in original). Indeed, proponents of CDA hold that its position as a tool to fight the oppression and contradictions present in everyday discourses, allows it to serve as a way to see through the layers of discourse which shape the world we inhabit and, as a result, our perceptions of it (van Dijk, 2006; Fairclough, 1992). Yet others in the field see this use of CDA as a potential weak point due to its heavy reliance on interdisciplinary research methods (Jones, 2004) and the circular analysis which informs many CDA interpretations (Stubbs, 1997).

It is here that defining CDA becomes rather difficult. In examining CDA critically, it can be seen as simultaneously existing inside and outside of linguistics. That is to say that—as in Wodak's discussion of what is meant by 'critical' (above)—there are many ways to conceptualize CDA and to put it into practice. While CDA has alternately been seen as a theory, approach, methodology and tool (discussed below) it is most often understood as being linguistic in nature. Nonetheless, for the same reason that there is a need for Wodak to clarify the meaning of the term 'critical,' there is leeway in how CDA research has been understood by researchers and how it has been carried out in research. To understand the sheer amplitude of the area itself as well as the ways that it has been employed in research over the years, it is first important to understand the underlying characteristics of CDA and how these tie into its methodological strengths and weaknesses.

1.1.1.1. Approach/Characteristics

Although CDA is often thought of and referred to as a 'theory' by many in the greater academic community, CDA researchers are quick to point out that CDA can be defined more accurately as a methodology, comprised of a set of interdisciplinary tools and approaches which are combined to analyze discourses through varied and shifting methodological strategies (see Weiss & Wodak, 2003; Wodak, 2000). Indeed, this is how it has principally been used in research despite any individual labels which have been placed on it. Regardless of the novel ways that CDA has been used (e.g. Ruth Wodak's DHA (2007)), it is—in essence—a way of using varied linguistics-based approaches to expose underlying discourses as they pertain to and are derived from political and social issues. As Fairclough (1992) points out, although CDA and other linguistic research methods focus on some similar themes, CDA additionally focuses on social theory

applications and examines macro-features in language use. Because of this, it is very closely tied to ideological concerns. This is one of the most vital parts of CDA research, however, it is also what has generated disaccord in regard to how CDA should be carried out and what its place is within linguistics.

In fact, in looking at some of the criticisms which have been presented over the years with regard to CDA it becomes clear that many of them have been based on CDA's reliance on the researcher's personal criteria. This is especially true in cases where the researcher's involvement in ideological components of CDA research have been called into question (Poole, 2010; Collins & Jones, 2006). Unfortunately for CDA researchers, though, it is quite difficult to separate ideology from methodological concerns due to the very nature of looking at discourse critically. As Fairclough (2003) points out, this separation is difficult because looking at discourses with critical eyes necessitates the presence of ideology, which Fairclough defines as, "…representations of aspects of the world which can be shown to contribute to establishing, maintaining and changing social relations of power, domination and exploitation," (p. 9). In utilizing this definition, the use of ideology as a lens through which to view research becomes problematic because it is not only connected to CDA methodology, but is a pillar of CDA research. Because of this it proves difficult to address ideological components of CDA research without having to address the whole of CDA.

What is noteworthy in looking at CDA research based on the connection between ideology and methodology is that, although there would appear to be little agreement within the field in many respects, there is a general consensus among researchers that the inclusion of ideological concerns is vital in effectively carrying out CDA research. For instance, van

Dijk (1997) portrays discourse in general as being inexorably tied to cognition and cognitive processes saying that, "…although it is sometimes useful to abstract from the mental nature of grammars, rules, norms, knowledge or opinions in an account of discourse and communication, it is obvious that a fully fledged theory of discourse would be seriously incomplete without a mental (cognitive or emotional) component," (p. 17). While this is a legitimate point concerning language—especially when it is examined within ideologically charged arenas—it is important to note that this sort of dependence on ideology as a component of CDA research is precisely what has been harshly criticized by outside researchers. In general, these criticisms have not come about due to the acknowledgement of ideology's place within the discourses being studied, but rather due to ideology's presence in the studies themselves as well as in researchers' interpretations of results (Poole, 2010; Haig, 2001). However, while these criticisms have been leveled against the stances of the very researchers working in the area, others have criticized the method as a whole (Jones & Collins, 2006; Jones, 2004). This is because CDA is seen to borrow so liberally from other fields that on one hand it can be seen to lose itself in methodological approaches, while on the other hand it can be seen to benefit from not being seen as wholly linguistic in nature (de Beaugrande, 1997). That is, CDA can be seen as very applicable since it is involved in so many areas, but very credible since it is presented as part of one area (linguistics). It is precisely because of these factors that a new approach is being discussed, tested, and proposed in the present study.

The interdisciplinary approach taken in CDA research is wholly necessary since CDA research scrutinizes social and political problems while viewing them as having an effect on discourse in society (Carvalho, 2008). Since CDA examines power and ideology as

manifested in social and political contexts (Fairclough, 1995), it is of vital importance that authors confront the underlying social and cultural factors that affect not only the construction of discourses, but their interpretation by and effect on the greater society. Additionally, its interdisciplinary bent allows CDA to be applied to virtually any venue in which language is prominently used, making it not only applicable, but adaptable (Carvalho, 2008).

Since, as the above discussion suggests, CDA is a more or less linguistically based approach which can be applied to virtually any practical issue within the purview of linguistics or social sciences, it is a very important tool in research in that it can form a sort of theoretical bridge between palpable, everyday instances of discourse (political language, for example) and linguistic theory. This is especially true when looking at the research carried out by authors such as Fairclough (1992) and Carvalho (2008). Both of whom have proposed a sort of cyclical interaction between discourse and what Carvalho calls 'Mediated Discourse.' This refers to the connection between the greater societal discourse's connection to 'reality' (in Faircloughian CDA) and society's interpretation and perception of discourse as related to 'reality' (in the case of Carvalho). What both of these authors have in common (at least in the investigations which were included here) is a focus on the end result of their research. While it could easily be argued that all research is focused on achieving an end result, what marks these studies as unique is that they seem to focus on the enormous influence which discourse has in society and not as much on the linguistic features within the discourse itself.

In the case of Carvalho (2008), for example, CDA is seen as an interconnecting group of approaches which can be used to view discourse's role within society. This can be seen

temporally, as in the cases of Wodak's DHA and Fairclough's Intertextuality (see Carvalho, 2008 for discussion) or by examining what Carvalho calls the 'life' of media representations in which, "understanding the evolution of matters such as war, terrorism or climate change, and the ways they are interdependent in relation to the media, is [seen as] one of the most important contributions to be made by social researchers," (p. 164). What is important to note in these different approaches to CDA, is that while they all view the same sorts of features within a given discourse, they focus primarily on the macro features of said discourse (i.e. ideological concerns, effects on society, social influences, etc.) while the explicitly linguistic (micro) features principally serve as a stage in justifying findings.

In this sense, van Dijk (2006) goes a step further in inferring that—based on his research regarding the presence of racism in discourses—in general, discourse contributes to the formation and persistence of certain ideologies. Due to this proposed intimate connection between general societal discourses and society itself, CDA necessarily employs a broad-based multi-disciplinary approach which permits researchers to draw conclusions based on as many linguistic and non-linguistic factors as possible. CDA as a whole—in this respect—is essentially based on the connection and interaction between discourse, society, and cognition at both individual and collective levels. It is important to note here that this basis for the field almost entirely excludes linguistics from its methodological approaches, findings, and conclusions. This is discussed in greater detail below.

Although CDA is a wide reaching methodology insofar as it employs techniques and approaches from a variety of areas, at its heart it has fairly simple objectives. In its most basic form, CDA is comparable to standard DA. However, due to its interdisciplinary

approach as well as its critical slant, CDA deals with the same micro elements of discourse as does DA while focusing more prominently on social phenomena in a bid to make 'hidden' discourse strategies transparent (Wodak & Meyer, 2009). While this accounts for what makes CDA functional as a stand-alone approach to linguistic research, it also represents CDA's greatest weakness; as a linguistics-based approach it is hard for authors to effectively make arguments in other areas, such as historical research (see Jones & Collins, 2006; Jones, 2004) while as a multidisciplinary approach it is hard to make sound linguistic arguments (see Poole, 2010). Further, the fact that no one 'owns' CDA—that is, that it serves as more of a tool than a theory, or at the very least a set of varied approaches each based in a separate theoretical foundation (see Wodak & Meyer, 2009)—has complicated the interpretation of its place in linguistics in that many things have been done under the guise of CDA, while there has been little in the way of establishing concrete procedures for carrying out CDA research; while this has led to an enormous quantity of studies carried out under the banner of CDA, it has yet to result in a concrete approach that can be used in a variety of situations to obtain comparable results.

1.1.1.2. Background And Shortcomings

The current project is centered on a perceived lack of empiricism and objectivity within CDA research as well as a traditionally poor application to linguistic research (according to many authors in the general research community). While it can be effectively argued that CDA research is not explicitly 'linguistic' research, it is applied to language issues with such frequency that it becomes difficult to separate the two. Thus, for the purposes of the current project, CDA is viewed as a means of conducting linguistic research. Because of this, the problems addressed here arise as a result of situations in which CDA is directly

applied to linguistic issues. There are, indeed, a number of problems in the CDA approach to linguistic investigation. Principal among these are the methodological and theoretical approaches taken by CDA researchers in selecting and analyzing texts. Many authors attribute this 'weakness' in CDA approaches to the very interdisciplinary characteristics that typify CDA as a methodology (Poole, 2010; Prentice, 2010). Oftentimes these criticisms are based on the underlying approaches and themes within a given CDA study. For example, CDA has been roundly criticized because of perceived weaknesses in its conceptualization and understanding of the application of historical methodology (see Collins & Jones, 2006; Jones, 2004). Additionally, the approach has come under attack because of its conceptualization of discourse's place in text interpretation (Jones & Collins, 2006) and because of the underlying ideologies which inform the CDA research process (Poole, 2010), as well as for a wide variety of other reasons related to the approach's methods and conclusions (see Haig, 2001, for discussion).

A common argument that all of these critiques of traditional CDA research share is that the aims of the field have become lost in the very multidisciplinary inclusion which characterizes it. However, as Prentice (2010) indicates, the shortcomings of CDA research are also related to what they characterize as 'arbitrary text selection,' paired with the approach's characteristic subjectivity (see Poole, 2010; Prentice, 2010; Collins & Jones, 2006; Jones & Collins, 2006; Jones, 2004) and theoretical and methodological weaknesses (see, e.g., Stubbs, 1997). Regardless of the varied perspectives found in these criticisms, they share the common idea that CDA is based on subjective, political motives. While this is more true of certain CDA investigations than others, it can be seen in the very discourse of CDA research such as Fairclough's (1995) assertion that the approach is a means to fight

oppression, Poole's assertion (2010—in critiquing Faircloughian CDA) that the ideology of a text has a greater impact on findings than does the method itself, or de Beaugrande's (1997) characterization of those whose discourses are analyzed by CDA as "… [CDA's] opponents, the snobbish guardians and glib homeworkers," (p. 45). Not only have these ideological characteristics within CDA characterized it and its development, but they have informed nearly all criticisms of the methodology to the present day.

The present research intends to address these issues by utilizing a more explicitly data-driven approach to research and an approach to DA grounded in theory. In order to accomplish this, a methodology founded in corpus linguistics will be used to collect textual data; and data will be analyzed using an approach based on semantic theory. These approaches, their connections, and their places within the current project are discussed below.

1.1.2. Corpus Linguistics

One of the newest approaches to conducting linguistic research and, indeed, one of the most frequently used methods in conducting modern CDA research is Corpus Linguistics (CL). CL is a widely applicable approach to linguistics research which utilizes computer programs to amass and analyze enormous collections of texts. Due to CL's current popularity in linguistics research, it is assumed that the reader is familiar with the field in general as well as both its theoretical and practical underpinnings. For those unfamiliar with CL, Biber, Conrad, and Reppen (1998) and Kennedy (1998) offer simple and comprehensive introductions to the field, its applications, and its utility in conducting linguistics research.

One of the principle advantages of corpus linguistics is that it can be tailored to practically any experimental situation. The corpora to be studied in a given investigation can either be chosen by the researcher (from pre-compiled corpora) or can be constructed according to the type of language use being studied. Following text selection and/or corpus construction, corpus programs are capable of singling out specific language features within texts, and allowing them to be easily searched, categorized, viewed and analyzed by researchers.

Corpora are comprised of texts which are chosen according to certain criteria, ideally the smallest set possible (Sinclair, 1991), in order to form a representative sample of language use within a certain area. A representative corpus, in this respect, does not only have to be representative of language use in a particular area, but must be representative of language use with regard to a specific topic. Because representativeness in a well-constructed corpus applies to these two areas of general discourse, the construction of a representative corpus affords the researcher two principal advantages. First, any data gleaned from the corpus (assuming that it was well-constructed) will present the researcher with a statistically representative glimpse of language use within a particular area; and second, data analyzed from the corpus will be representative of the language generally employed in discourses regarding a specific topic. This thematic and linguistic representativeness allows corpus research to focus on language use characteristics in a measured and practical manner while discussing conclusions pertaining to it.

In the current research, for example, the corpora were representative of newspaper writing published online (linguistic arena) and language surrounding the ‘drug-related violence’ in Mexico and the United States (topic). Because of the criteria (discussed in

Chapter 3, p. 52) established for corpora construction, these corpora can be seen to be representative of newspaper stories, published online, reporting on 'drug-related violence' in Mexico and the United States[2]. When paired with the use of semantic theory, these corpora can be effectively analyzed on a more objective level than has been done previously.

1.1.3. Semantics

Semantics is a branch of theoretical linguistics (along with syntax and phonetics) which examines the meanings of words and the individual lexical features that make up and affect language. Every lexical item can be seen as a set of features which characterize it and, as such, semantic theory can be used to determine the basest (prototypical) meaning of a given word as well as how lexical items (LIs) contribute to overall linguistic meaning whether within a word, a phrase or a discourse. The first publication the discussion that would eventually become semantic theory was Katz-Fodor's discussion of semantics in 1963, which was a response to Chomsky's proposition that syntax was the foundation of language. The analysis of semantic-based linguistic thought offered by Katz and Fodor served as an answer to Chomskian generative grammar in that the authors proposed that, "…in no sense of meaning does the structural description which the grammar assigns to a sentence specify either the meaning of the sentence or the meaning of its parts," (p. 173). That is to say that the meaning present in natural language is a product of the presence or absence of the individual LIs which make it up. In this sense, semantic theory is

[2] 'Drug-related violence' is used throughout the present study to refer to a myriad of inter-connected issues related to the violence which has come as a result of altercations between the Mexican and US governments and drug-traffickers in Mexico since 2006. Although some texts included in the corpora report on civilian matters, all of them were connected to the violence currently occurring in parts of Mexico and the US, resultant of the drug trade in both countries (see Section 3.1.1.2, p. 64).

complimentary to many aspects of CDA in that it allows for a focus on the underlying meaning of a text through the use of verifiable linguistic theory.

As de Swart (1998) describes it, semantics looks at the linking of the *form* and the *content* of natural language. This conceptualization of semantics allows for the fact that competent speakers can both communicate ideas and understand language—as opposed to only one or the other. The difficulty in accurately delimiting the reach of semantic theory, as well as the dual importance of both 'communication' and 'understanding,' is made abundantly clear in viewing something as simple and common as paraphrasing, in which two distinct linguistic representations carry the same meaning based on the semantic uses which each features despite any surface syntactic differences. This balance between form and content described by de Swart is often seen as a split between semantic theory on the one hand, which posits that in saying anything there is an implicit meaning; and pragmatic theory on the other hand, which posits that the implicit meaning of a word can and is changed regularly according to a variety of linguistic, social, and contextual factors. An important factor in clarifying any confusion in this respect is that the *choice* of words may indeed hold a message, but that the words themselves are built upon a consistent and analyzable meaning which can in turn affect other parts of a given utterance (such as in the example of paraphrasing).

SP is, in many ways, an approach which responds to this understanding of semantics' place in language by allowing for semantic analysis while still accepting underlying meaning and intended meanings of LIs used, manipulated, and avoided on the part of the author. In the case of semantic prosody, this is accomplished through the analysis of connotation. This is possible because the semantic meaning of a lexical item is, as Morley

and Partington (2009) put it, part of a word's "DNA." Set against pragmatics—which holds a similar understanding of intended meaning's place in language—semantic prosody puts forth the idea that engrained connotations in LIs account for the shared understanding of utterances among humans. That is, where pragmatics relies on interpretation to understand a given utterance, SP proposes that underlying meanings are part of a speaker's grammar (for a complete discussion of this distinction and related issues, see Morley & Partington, 2009; Partington, 2007; Partington, 2004).

1.1.3.1. Semantic Prosody

The use of a semantic prosody-based approach in the current project will allow for a focus on individual words within discourse and their connections to surrounding words, individual texts, and the corpora as wholes. As a part of the Methodological Fusion upon which the current project is based, semantic prosody will allow for broader and more objective conclusions than have other, similar approaches such as that taken by Baker et al. (2008). In fact, the use of semantic prosody in the current project combined with the corpora being used effectively utilizes the strengths found in the Baker et al. study as well as Salama's (2011) study. That is, the present investigation uses a broader sample for analysis (as in Baker et al.) while using semantic analysis techniques in order to improve analysis (as in Salama).

This added objectivity comes from two features in the analysis. First, as some authors have pointed out, semantics is a theoretical part of linguistics which equates a scientific approach to understanding meaning and its interaction with syntax and other linguistic areas (see, e.g., Palmer, 1976; Leech, 1981); and second, the use of corpus analysis makes

it possible to apply the understanding of semantic meaning to a broader area of textual discourse. This is made more effective in that corpus linguistics and semantic prosody analysis are inextricably linked through their necessary methodological steps.

In fact, semantic prosody has its roots in corpus linguistic studies carried out by John Sinclair in the late 1980s. It was Sinclair who first put forth the idea that certain LIs occurred together time after time across many corpora and that these habitual collocations were analyzable. This consistent lexical behavior has been found in many linguistic areas over the years including in collocation, which led Sinclair to the investigation of schemata (Zhang, 2010) something that has since been expanded and altered to include what Louw (1993, as cited in Whitsitt, 2005) dubbed 'Semantic Prosody.' This initial foray into what would become semantic prosody was carried out through the semantic examination of irony. Although irony is still occasionally examined through SP (e.g. Partington, 2007), the field has expanded to include many other applications in more practical realms, such as studies of lexicographic applications (Ping-Fang & Jing-Chun, 2009) and the study of metaphor and idiom use in texts (Oster, 2010; Fillmore, Kay, & O'Connor, 1988). Nonetheless, despite promising applications, SP has been predominantly discussed in highly theoretical terms. That is, very few studies have been carried out using it as a prominent component of their methodologies. The majority of SP investigations have focused on showing SP characteristics in specific examples (occasionally as small as one word) looked at throughout mega corpora. Although this has brought forth interesting findings and could one day lead to a more complete understanding of semantics, in many ways SP has been underutilized in terms of practical applications. The method being

proposed here would utilize basic forms of SP analysis as part of a separate methodology with the end goal of improving upon existing approaches to CDA research.

1.2. New Method

The following section analyzes various approaches to CDA research that have been taken in the past. Additionally, it features a discussion of related methodologies and areas such as media studies and frame studies. Based on the discussion of these diverse methods, an argument will be made for the necessity of a new approach which synthesizes features from many other CDA-based methodologies.

A modified approach to CDA research is an important first step in addressing many of the criticisms which have been directed at traditional CDA since its inception. It is not claimed that the new approach presented here will completely resolve any of the existing problems immediately, nor is it claimed that a new approach will be without similar theoretical, ideological or methodological weak points. However, an approach which makes use of methods chosen with the plain intention of achieving methodological objectivity and applied with the same intention, could serve to move the field forward. For these same reasons, the print media (in this case, newspapers) can be successfully utilized as a basis for investigation. This is the case because of the general characteristics with which print media is—in theory—endowed. An objective approach to studying media discourses could effectively single out instances of ‘non-characteristic’ media discourse (i.e. instances of bias or ideological posturing). This is particularly pertinent in that the news media is founded on the principle of being non-biased and that the CDA approach is

often focused on finding and highlighting bias or underlying discourse patterns which push a point of view.

Aside from the fact that the news media is ideally objective, it is also a model medium in which to carry out discourse research for other reasons, not least of which is its connection with and availability to the public, something that is immensely important in CDA research due to CDA's being traditionally seen as a sort of tool for informing the masses. Not only does the mass media represent the main outlet through which the general public gets its information about the world; but, as many CDA researchers have pointed out (Gutiérrez Vidrio, 2010; van Hout and MacGilchrist, 2010; Carvalho, 2008), media and policy have a cyclical relationship in which policies are reported in the media, influencing people's opinions of them and, in turn, affecting future policy-making by way of public opinion.

Because of this cyclical interaction with peoples' lived realities, the media has a tremendous effect not just on people's lives and their perceptions of the world, but on the very world which they perceive and in which the media exists. Due to this connection between social and political issues and discourse, a traditional CDA study is neither overtly nor intentionally subjective; however, the fact that subjective, ideologically based interpretation on the part of the researcher comprises one of the most important analysis techniques in CDA research often results in a strong ideological discourse being implanted in CDA studies. Jones (2004), for example—making particular reference to Norman Fairclough's place in CDA research—criticizes CDA's 'explicit political agenda,' claiming that the discipline and its corresponding goals are obscured by its inherent ties to certain political ideologies and approaches to research.

Despite these well-documented short-comings, many CDA findings are well-presented and, by all outward appearances, reliable in their conclusions. Because of this seeming-credibility within traditional CDA investigations, a more positivist-based approach to CDA questions (in this case, using a combination of corpus linguistics and semantic theory) would be *at least* as methodologically and theoretically sound as traditional approaches. What this means is that the approach being proposed here has the potential to offer an innovative way to carry out CDA-related research while simultaneously avoiding many of the problems which have been identified in past applications of CDA; however, even if the current project finds nothing new with respect to the effectiveness of CDA methodologies, it will still provide a more linguistically centered and methodologically objective approach which can be utilized in future investigations.

1.2.1. Current Study

As a result of the criticisms which have been raised against CDA research (briefly outlined above), the present research was designed to respond in the simplest and most effective way possible. This was accomplished by creating a methodology that addressed the most common individual criticisms of CDA in a single methodology. In order to accomplish this, the current project was designed to be a combination of Corpus Assisted Discourse Analysis (CADS) (Freake et al., 2011) and SP. This methodological combination was intended to be applicable as a methodological tool for carrying out CDA-based studies. Essentially, the study is based on SP approaches and Methodological Synergy (Baker et al., 2008).

Methodological Synergy is an approach to CDA which was first proposed by Baker et al. (2008). The proposed methodological fusion is broadly based on the combination of CDA and Corpus Linguistics. Since Baker et al.'s initial experimentation in applying Methodological Synergy[3], other authors have begun to embrace the approach as a way to draw more prominent, empirically provable and applicable findings from the study of CDA issues (see Freake et al., 2011; Salama, 2011). Nonetheless, the Methodological Synergy approach still lacks in key areas. As Baker et al. (2008) point out, CDA never was necessarily a method in and of itself; in fact, CDA researchers utilize any method available to them so long as it proves complimentary to the study being undertaken. Further, CDA has traditionally relied heavily on qualitative data. Because of this, the field has come under attack for being weak in relation to linguistics (Orpin, 2005). Authors who have criticized CDA in this respect have principally done so in relation to CDA's heavy reliance on the ideologically based interpretation of findings. Despite the fact that this is a strong argument against CDA, it is primarily true of research which relies *exclusively* on CDA methodology. In the Orpin study (cited above), for example, CDA was only employed in examining the ideological components of the findings. With this sort of methodological balance in mind, many authors have tailored their approaches to account for criticisms of CDA. One of the most obvious attempts at addressing these shortcomings is Methodological Synergy (Baker et al., 2008); however, although Baker et al.'s approach serves to place more importance on empiricism within CDA research, much could still be done to make the approach more objective as well as more tangibly based in linguistic

[3] Throughout this study the terms CADS and Methodological Synergy are used in reference to the combination of CDA-related methodologies and corpus linguistics. The current study is essentially a combination of both (CADS is a broader methodological approach and Methodological Synergy utilizes a mega corpus—something not present here). However, a large portion of the research currently available in the area is based on Methodological Synergy and, thus, the term is used to refer to the CDA-based methodology used here.

theory. These methodological problems are not solely relegated to the Baker et al. approach to 'Methodological Synergy,' though. One of the only papers published to date which prominently features the use of Methodological Synergy has been Salama's (2011) study of the contrasting use of 'Wahhabi' and 'Wahhab' in two texts. This study has likely come closer than any other to combining semantics with Methodological Synergy. Nonetheless, this study could still be improved upon and made more practical and objective in relation to its methodological design (discussed in Chapter 2, p. 29).

With these approaches, studies, criticisms, and findings in mind, the current project aims to examine textual discourse while avoiding many of the pit-falls encountered in traditional CDA research and while building on the foundation that was laid by Baker et al. (2008) with their Methodological Synergy approach. This was accomplished on two fronts. First, the present study was carried out on a stratified random sample of newspapers from two countries (the United States and Mexico). This has not been done extensively in the field to date. Most of the studies published in the area have focused on the use of mega corpora (or at the very least have employed them in order to draw comparisons) and have principally looked at English. Second, SP was utilized when analyzing the results of the corpus analysis. In this way, statistically salient LIs and their collocates (in both English and in Spanish) were looked at in order to examine contrasting, parallel or distinct prosodic features (elaborated in Chapter 3, p. 52). This contrasting analysis was intended to single out any underlying discourse tendencies which may have been present in the corpora, as well as to indicate how the general discourse was different or similar across two corpora of texts based around the same issues.

The analysis which was carried out here was undertaken with the intention of exploring the hypothesis that it is possible to carry out CDA research using a methodological approach which eliminates the necessity for predominately subjective analysis of and commentary on the texts analyzed. In order to test this hypothesis, the current project was designed around the following four fundamental assumptions:

a) Semantic prosody can be used as a research tool in examining corpora written in different languages (in this case, Spanish and English).
b) Semantic prosody can be applied to the analysis of texts from distinct linguistic communities (Mexico and the United States) in order to examine similarities and differences in the predominant media discourses of each country.
c) Semantic prosody theory can be applied to corpus driven CDA (specifically to Baker et al.'s (2008) Methodological Synergy approach).
d) Semantic prosody and corpus linguistics can prove to be effective tools in carrying out CDA research and can be applied to future CDA research projects; thus eliminating a large portion of subjective guess-work in the area.

Together, these points served as the impetus for proposing, designing and carrying out the present study. While the overall goal of the study was to put forth a more complete way of addressing methodological weaknesses in CDA research, importance was also placed on determining whether or not the use of SP was a feasible way to accomplish this goal. Obviously, if it were impossible to apply any or all of the approaches listed above it would not be feasible to attempt the current project. Nonetheless, there have been enough applications of the individual approaches as well as various combinations of them (see

Chapter 2, p. 29) that the use of SP as a complement to CDA-research was seen as a feasible next step in carrying out this sort of research.

While the above expectations informed the underlying processes utilized in designing and applying the methodology used here, the research process itself was based on the following three research questions. These questions were created for the purpose of being able to determine the strengths and weaknesses of the current project and in what ways the results of the current project could be applied to future research within the field. The research questions are presented below and discussed thereafter.

a) Is an analysis based on the combination of aspects of Methodological Synergy (Baker et al., 2008) and Semantic Prosody capable of analyzing discourses present in two separate languages?
b) Can the combination of certain aspects of Methodological Synergy and Semantic Prosody result in a more comprehensive and less subjective approach to the goals of traditional CDA research?
c) Are the methodological steps utilized here applicable to future studies in CDA?

These research questions formed the pillars of the present study because they set the tone for what the rest of the research process would address. In the most general sense, the present research was focused on determining the feasibility of utilizing a combined methodology for approaching CDA research. However, within this it was also important that the framework used here be relevant to the traditional goals of CDA. That is, the methodology used here was intended to be a complementary approach to traditional CDA and not an entirely separate methodology. With this in mind, it was also vital that the

current project be applicable to future studies, thus moving the field forward. In the interest of laying out the need for the current project as well as the reasoning behind its design, the following chapter presents the concepts and methodologies which informed the design, proposal and execution of the present research.

2. Literature Review

In order to carry out traditional CDA research from a novel perspective, while additionally proposing and testing a new method through which to carry out future CDA-based investigations, it is first important to discuss the field itself. As a result, the following section highlights the history of CDA, its development, and some of its applications within the literature. Further, this section discusses the problems which have been found in the traditional CDA approach to linguistics research as well as the linguistic research methods which were used in the current study as potential solutions to pre-existing methodological weaknesses.

With the intention of laying the groundwork for the investigation which was undertaken here, this section reviews past and present research in all of the fields included in the present methodology. This is important not only in clarifying the logical steps which have influenced the current research but in presenting the methodologies being combined here in terms of how they can serve to solve some common problems found in traditional CDA research. As such, all of the research methods discussed below are looked at with a basis in their connections (both explicit and implicit) to CDA research as well as in their historical presence and achievements in various fields of linguistics research.

2.1. Critical Discourse Analysis

CDA—as was discussed in the introduction (see page 1)— is an approach to linguistics research which differs from classical DA, "…in not just describing discursive practices, but also showing how discourse is shaped by relations of power and ideologies, and the constructive effects discourse has upon social identities, social relations and systems of

knowledge and belief, neither of which is normally apparent to discourse participants," (Fairclough, 1992, p. 12). Although CDA has been applied to a great number of issues within linguistics research, all while using many different methodological approaches, it is safe to say that Fairclough's description encapsulates the basest characteristics which set CDA apart from other areas of linguistics (most notably DA) and also what has caused there to be such debate in the field over what CDA is, where it is going and whether it plays a legitimate role in advancing linguistics research as a whole.

One of the most difficult aspects of discussing CDA is actually pinning the area down in concrete terms. The area, or methodology, or theory (depending on the author consulted[4]) can have a number of uses, from political speeches, to racism, to advertising; and not surprisingly, the history of CDA is similarly murky and dependent on the perspective of each author. What is generally agreed upon in the field is that CDA came about sometime between the mid-1970s and early 1980s as a response to both DA, which came to prominence in the 1960s (van Dijk, 1997), and critical linguistics (Wodak & Meyer, 2009; Fairclough, 1995).

In its simplest form, CDA research can be conceptualized as a pyramid representing the multi-directional interaction between discourse, cognition, and society (van Dijk, 1997). In this model, discourse is seen to be a socially and ideologically based cycle where ideologies are seen as, "…representations of aspects of the world which can be shown to contribute to establishing, maintaining and changing social relations of power, domination and exploitation," (Fairclough, 2003, p. 9). Due to this perceived connection between

[4] In this work, CDA is referred to as an *area* or *field* (in reference to the body of work of authors working in CDA) as well as either an *approach* or a *methodology* (referring to CDA's use in carrying out a given study or set of studies).

discourse, ‘power,’ ‘domination’ and ‘exploitation,’ researchers’ ideological leanings and understanding of modern and historical events become important in interpreting discourses as well as in selecting data for studies. For example, in looking at political discourses one must account for political knowledge, underlying political strategies and ideological stances of the originator of the discourse being examined, as well as numerous sociological and cultural factors in order to render a suitable analysis which addresses discourse, cognition, and society (van Dijk, 1997). In order to effectively address these parts of the discourse, it is necessary to utilize a methodology which is focused on looking at language use through the lens of ideological and social factors, and, indeed, this is what has been done in much of the published CDA literature to date. Despite differences of opinion in the area, the presence of ‘discourse, cognition and society’ can be seen in most studies. In Wodak and Weiss’ (2003) four-level approach, for instance, CDA research is seen as a way to ‘triangulate’ discourses in context using four steps. First, the language or text is looked at descriptively. This is followed by an examination of the relationships between the ‘utterances, texts, genres and discourses’ found therein. Finally extralinguisitic (socially based) variables are examined and these (as well as the discourse itself) are looked at in terms of ‘sociopolitcal and historical contexts.’ The presence of the ‘pyramid’ is also found in other, less well-known approaches and forms the basis for nearly all CDA research.

It is because of the near-universality of the two underlying characteristics present in the models mentioned above (ideology and social factors) that there is a prevalence of interdisciplinary methodologies within most studies in the area. Consequently, these characteristics also contribute to many of the criticisms leveled against CDA, primarily that studies fall prey to ‘circularity’ and that they are often not based on ‘firm linguistic

evidence' (Stubbs, 1997). If discourse is seen to affect not only individuals but also society and—as a consequence—future discourses, then understanding discourse and effectively utilizing CDA methods necessarily requires the use of varied discourse arenas. This is true methodologically speaking in that authors may borrow from other fields while conducting research; however, it also signifies that CDA is—for good or ill—one of the most flexible areas of linguistics research in terms of applicability. Because 'discourse' is present in so many areas of human existence and is, as a result, difficult to aptly define (see van Dijk, 1997), CDA can be carried out in many different linguistic areas. Since its inception, CDA has focused on a number of issues, while, "gender issues, issues of racism, media discourses, political discourses, organizational discourses or dimensions of identity research have become very prominent. The methodologies differ greatly in all these studies on account of the aims of the research and also the methodologies applied: small qualitative case studies can be found as well as large data corpora, drawn from fieldwork and ethnographic research," (Weiss & Wodak, 2003, p. 12). While there is a great deal of diversity in the applications that have been realized by CDA researchers, because of its reliance on ideology and social matters, much of the area's attention has been devoted to examining political and media discourses. The current study picks up where many CDA examinations of the media have left off by addressing methodological weaknesses and focusing on the media's presence in modern society.

2.2. CDA Studies of Media

In order to carry out the current research it is first important to examine CDA research as it is tied to studies of the mass media. This is the case in that the current project is an examination of media discourse using CDA as a starting point. Further, although there

have been media-centered studies using CDA in the past (see below) much of the published research on media language use has existed as a fundamentally separate entity from CDA. With this in mind, the first step in approaching a CDA-centric media investigation was to highlight the shared characteristics and complimentary methods between CDA research and general media studies.

Although the media is an enormous and prominent source of socially significant discourse and language use, the linguistic study of media is relatively undeveloped. Much of what is today thought of as Media Studies was initiated by the Glasgow Media Group under the direction of Greg Philo in the mid-1970s (Glasgow Media Group, 2010; Fürsich, 2009). Following initial forays into the study of the media's presence in society, the discipline began to grow, finally expanding to form a large, extremely interdisciplinary branch of DA. Eventually, with the introduction of research components related to Critical Linguistics in the late 1980s, Media Studies began to have a presence in CDA-based research as well. In this respect, CDA came to be used as a means to explore the symbiotic relationship between media and discourse construction. This understanding of discourse's place in the media relies on the idea that a society or culture's day-to-day reality influences what is presented in the media which, in turn, influences the same society or culture's day-to-day reality (Gutiérrez Vidrio, 2010). This is encapsulated quite well in Carvalho's (2008) concept of Mediated Discourse.

The combination of Media Studies and CDA came about as a natural progression due to CDA's simultaneous embrace of aspects of sociology and political science, as well as syntax and semantics. Due to CDA's focus on the discursive presence of divisive social events in modern society and the news media's reporting on such events, the combination

of CDA and Media Studies provides a doubly-representative view of modern critical linguistic issues. While CDA highlights the presence of discourse in media language and the potential consequences, Media Studies serves as a window into the prevalence of such language use in practical areas. Indeed, for these same reasons, numerous worthwhile studies have come out of critical analyses of media. Many of the studies carried out involving CDA as applied to Media Studies have focused on 'framing' (Túñez & Guevara, 2009, for example) and similar instances of what authors have called 'imaging' and 'portrayal' (see, e.g., Berger, 2009). These studies are important examples of the successful pairing of Media Studies and CDA because they show not only the underlying discourses present in language but common occurrences of such discourses.

When viewing the use of combined CDA and Media Studies methodologies, 'framing' studies would seem to be some of the most common approaches. These studies are based on the concept of discursive frames, which according to Lewis and Reese (2009) are,

> a central facet of political communication, frames define the terms of debate; shape public opinion through the persuasive use of symbols; and, when most effective, lead to public policy change. They serve as the primary vehicle through which public officials, the news media, and other elites exercise political influence over one another and the public at large. (p. 85)

According to this conceptualization, frames carry out many of the same functions that CDA is charged with studying in relation to discourse; this is primarily true with regard to the idea that frames can affect 'change' in the public and private sectors. Because of this, framing studies represent one of the most common approaches within the discipline.

Because frame research in media language is so conceptually similar to the fundamental principles of CDA, the topics in which framing is examined are usually tied to socially important, controversial, or political themes. Lewis and Reese (2009), for example, studied framing in the media following the events of September 11th in the United States. The research team viewed the phrase 'War on Terror' as a frame which, to them, constituted a, "…socially shared organizing principle…" (p. 86). Their study examined this term's use and rise to prominence within journalistic texts. In order to gain a perspective on its use as well as journalists' relation to the framing of 'War on Terror,' Lewis and Reese carried out interviews with journalists in order to delineate journalism's role in a similar way to that used in Carvalho's (2008) Mediated Discourse. That is, a cycle in which journalists put forth a term (through publication) with a certain perspective—influenced by their day-to-day realities. The cycle which Carvalho proposes—and which influenced the current proposal of Methodological Fusion—involves a given term being taken up by the press (usually after first being emitted by another entity) with a certain context/perspective in mind; and finally the term is disseminated to the public who gleans a certain meaning from it. This meaning then goes on to influence further media coverage involving the term. Carvalho posits that this is true in that the connection between discourse and society means that, "each discursive event is dialectically tied to society insofar as it both constitutes and is constituted by social phenomena," (p. 162). Versions of this sort of cycle are present in many areas of the published literature, and indeed form the very foundation of traditional CDA research. Oster (2010) (citing Kövecses, 1990) stresses that even something as complex as emotion can be described through lexical analysis in that it is present in language use through the use of metaphor and collocations (both of which are obviously tied to 'social phenomena' as well).

In one of the most transparent examples of Carvalho's (2008) process, the phrase 'War on Terror' was first put forth by politicians before being spread and given currency through its presence in news media where it has since become a very commonplace concept among the public (Lewis & Reese, 2009). In Carvalho's model, after a term gains currency through media permeation, the public's perspective progresses to the point that it affects policy at which point (in one way or another) the cycle begins anew. This can be seen in the same example of 'War on Terror' where the term went from politicians to journalists to the public before becoming a divisive ideological stance in American politics which has since entered the common lexicon and influenced numerous public and private policies.

Many other authors have used similar strategies to identify and understand framing in media discourse. Some of these studies have taken a more straightforward approach to the field, whereas others have viewed framing from a more abstract perspective than in studies like that carried out by Lewis and Reese (2009). Many of the studies which fall into the more theoretically concrete end of the spectrum have addressed 'imaging' and 'portrayal' in the media. Berger (2009), for instance, examined the palpable effects of media framing on a population in relation to disaster coverage in the media. By analyzing what Berger identified as the juxtaposition between the 'mediated representation' and the 'lived reality' of New Orleans following hurricane Katrina, the author examined the, "…merger of journalistic crisis and state breakdown [which] involved, among other things, mutually reinforcing notions of a crime-and-punishment spectacle that demonized much of New Orleans' population that weathered the storm," (p. 491). Thus, the Berger study gives credence to Carvalho's (2008) concept of Mediated Discourse by showing two sides of it. Berger's conceptualization of a 'representation' and a 'reality' within news coverage

corresponds with Carvalho's cycle in which the media portrays reality through reporting which is presented as unbiased and representative; this, in turn, influences public perception and policy.

In a similar vein, Davis and French (2008) looked at the connection between news and day-to-day life among New Orleans citizens following Hurricane Katrina. Focusing particularly on media characterizations of New Orleans citizens, Davis and French concluded that, "Post-Katrina news depictions were found to rely on specific rhetorical devices and semantic strategies, as well as seemingly value-free terminology (e.g., victim and survivor) to publicly discuss certain citizens involved in the event. However, such discursive construction had the rhetorical result of shifting blame onto those citizens," (p. 244). The difference between this study and the Berger (2009) investigation is that instead of viewing coverage as a means of portraying a different lived reality (as Berger did), Davis and French viewed framing in news coverage following Hurricane Katrina as a reflection of pre-existing frames regarding New Orleans. This is similar to the concept of Mediated Discourse (Carvalho, 2008) in that discourse's influence on society and life is viewed as a sort of cycle in which pre-existing discourses are important in shaping media discourses, which in turn influence lived reality and the perception of it. In the case of Davis and French's study, the frames examined essentially consisted of cultural stereotypes based around what the authors termed New Orleans' reputation as a city that is 'free-wheeling' and 'easy going.'

While the above studies focused on concrete examples of rather isolated uses of discourse in the media, other prominent investigations have focused on the more nebulous, philosophical side of media discourse studies by looking at individual fragments of news media language use as they relate to framing. These studies have included examinations of

media characteristics such as framing in leads (the opening sentences in newspaper articles) in internet news articles (Zillmann, Chen, Knobloch, and Callison, 2004) and the historical development of individual lexical items (LIs) within the media (see Peñaranda Contreras, 2010) according to what Berger (2009) calls 'lived reality.' Despite the fact that investigation of historical discourse transformation is quite common (see Peñaranda Contreras, 2010; Wodak, 2007; Törrönen, 2004) when analyzing the formulation and propagation of frames a less ample gamut of research approaches is present in the literature.

Nonetheless, researchers have looked at both the construction of frames and the development of news media reports. This has been done on both macro and micro levels. Túñez Guevara (2009), for example, looked at the presentation of one story through the print news media of four different locations (Mexico, Portugal, Spain, and Galicia). This approach allowed for the examination of contrasts between Mediated Discourses according to the media environment in which they were present. What was most noteworthy in this study was the way in which it was approached. By comparing media discourses according to country of publication—while controlling for topic—Túñez Guevara was able to single out differences and see them as something more significant than just the perception of discourse characteristics. That is, while a difference within a single country could be indicative of differing opinions or perspectives, international differences (and, by the same token, similarities) could be seen as indicators of differences in presentation and interpretation. This sort of control over the analysis of media discourse is what informed the methodological approach designed for the current study. If variables can be controlled well enough, and if the language present in the texts examined is 'neutral' enough, then—in theory—discourse characteristics will 'rise to the top' based on their mere presence in the texts analyzed.

This same sort of methodologically controlled analysis has also been carried out on a micro level, Van Hout and MacGilchrist (2010) 'followed' a news story through the publication process in order to observe and analyze, "…how discursive transformations shed light on journalists' writing practices… [they also] argue that the notion of framing should be decoupled from the intent to mislead; [because] every news story, as indeed every type of linguistic expression, must *per se* be 'framed,'" (p. 170). What this means is that Carvalho's (2008) concept of Mediated Discourse has such an effect on discourse formation and presentation that it becomes impossible to encounter media discourse that is not influenced by social and political factors in some way. This outlook on media discourse has been supported through other examinations of media language use based on Wodak's DHA (see Peñaranda Contreras, 2010; Törrönen, 2004).

These sorts of studies are not as common in the literature as are frame-based studies. Nonetheless, in the same way that studies like Van Hout and MacGilchrist (2010) have examined the micro-development of frames, authors such as Peñaranda Contreras (2010) and Törrönen (2004) have analyzed the macro-presence of discourse constructions as well as the changes that these constructions incurred over time; in the case of the Peñaranda Contreras study, the research centered on the changing use of lexical forms in news reporting during the 1960s and 1970s. The study was carried out using Colombian newspapers from two decades which were analyzed for the way in which they referred to those involved in the drug trade. The study found that over time, due to changes in day-to-day life in the country as well as a general shift in Colombia's lived reality, newspapers shifted their use from the term 'marihuaneros' (loosely translated by the author as *drug fiends*—employing American colloquial speech, one could even go so far as to suggest *pot heads*) to 'mafiosos' (translated by the author as *drug lords*). Similarly, Törrönen (2004)

analyzed articles about drugs and drug use viewing the rhetoric employed therein through the political context in which it occurred. Törrönen found similar changes according to the overall public and political opinion of the issue at a given time. In both of these studies, the most important binding characteristic is their focus on discourse and Mediated Discourse (Carvalho, 2008). Although both studies utilized methodologies similar to those present in some of the preceding discussion, the brunt of their arguments seems to be that discourse can be seen in a society's day-to-day 'reality' and that a society's day-to-day 'reality' can similarly be seen in discourse use.

The one negative characteristic which these studies share (at least in terms of the CDA weaknesses addressed in the present work), is a basis in the liberal selection of texts and parts of text for analysis. This methodological characteristic is certainly not relegated to the Peñaranda Contreras (2010) and Törrönen (2004) studies, either. Indeed, it is one of the most prevalent 'weaknesses' in the whole of CDA research. This characteristic is problematic in the field because it presents what Koller and Mautner (2004, p. 218, as cited in Baker et al., 2008) call a 'hidden danger' because, "…the reason why the texts concerned [in a given analysis] are singled out for analysis in the first place is that they are not typical, but in fact quite unusual instances which have aroused the analyst's attention." Because of its prevalence in the published research as well as the glut of critiques regarding it, the 'liberal' selection of texts is one of the first aspects of CDA research to have been effectively addressed by means of an alternate methodological approach.

2.3. Corpora and CDA

One factor which all of the studies mentioned above have in common is their basis in traditional CDA. That is, they look at different aspects of public discourse, but all share very specific methodological characteristics both in their gathering of texts and in their

analysis of those texts. Despite the fact that all CDA analyses are conducted using a corpus in the traditional sense (that is, a principled collection of texts), until recently there have been very few which have made use of electronic corpus analysis programs in their research. As was discussed above, the use of electronic corpus tools is more prevalent than ever. This is not only because of the sampling ease which these programs provide the researcher, but because electronic corpora provide a statistically based, verifiable, and—above all—consistent and objective research tool to investigators. Despite the advantages of using electronic corpus tools, the use of such programs has only recently become commonplace in CDA research.

In fact, the first prominent large-scale instance of pairing of CDA and corpus linguistics in a study was the Methodological Synergy approach, introduced by Baker et al. (2008). Although there have been other studies in the past which paved the way for the Methodological Synergy approach (see Baker et al., for discussion), this approach was unique in that it proposed the use of electronic corpus analysis programs as a means of combining corpus linguistics and CDA with the finality of demonstrating that the two approaches (CDA and Critical Linguistics) could be complimentary tools in carrying out research. Because this was one of the first attempts at marrying the two methods, the Baker et al. study—like the current study—was exploratory in nature. The study focused on the discourse present in printed news in England. The study was conducted using a 140 million-word corpus of news reports about British refugees and asylum seekers. Being as the study focused more on exploring a new approach than on the analysis itself, the main offering of the Baker et al. research was the conclusion that such studies are feasible and can potentially aid in the expansion of the field. The study did just that. As such, the most

important feature in the Baker et al. study was an overview of how such a methodology could be employed and how the results could be measured and analyzed by researchers.

Following the publication of the Baker et al. (2008) paper, one other author appears in the literature as having utilized their recommendations (at least under the same banner). Salama (2011) used the Methodological Synergy approach to analyze and interpret collocations in texts dealing with Wahhabi-Saudi Islam. Salama selected key words from two books regarding the subject and conducted a CDA-based collocation analysis in order to examine frequent (positive and negative) lexical uses within two corpora. Salama's research examined these concordances and used them to draw statistical conclusions about the uses of 'Wahhabi/Wahhab' within the texts examined. Based on the co-occurrence between selected node words and their collocates within the corpus, Salama was able to draw conclusions about LIs' use and significance within the texts; thus shedding light on underlying ideological motives within the corpora.

In Salama's (2011) study, the main focus of analysis was on the use of collocates in relation to paradigms. That is, the examination of frequent lexical occurrences as they relate/contribute to overriding views within a society. The study was carried out using two books regarding Wahhabi-Saudi Islam and was based on the analysis of the collocations of statistically prominent node words using Key Word in Context (KWIC) analysis. Despite these methodological measures, the research was still based on ideological interpretation in that both texts which were examined were selected due to their portrayal of a certain type of ideology (they were chosen due to the presence of antagonism in both texts). While Salama does compare language use in the corpora to a general corpus of American English (the *Corpus of Contemporary American English*) (a solution to many methodological problems

in CDA, proposed by Stubbs in 1997), the analysis is based on the antagonistic language used within the two books examined. As a result of this, the research findings highlight the very linguistic antagonism in both books which was acknowledged from the start of the research. In this respect, the Salama study serves as an exploration of the use of SP in corpus assisted CDA research. Also, the SP-based methodological tools Salama does use, are deployed in similar ways to many past studies. That is, node words were selected and were then searched with the intention of delineating their overall prosody within the corpora. Because of this, Salama's results—though immensely helpful in informing the methodology used in the present study—must be viewed critically in that the research effectively highlighted the antagonistic language which was already recognized in one way or another in beginning the research, and which was deemed a notable feature in the corpora from the beginning. Nonetheless, the Salama study is one of the most vital studies published to date in regard to the current project. While it can be criticized for focusing on antagonistic language within antagonistic corpora, it is one of the only prominent articles (at least at the time of this writing) which has taken the extra step—being added to the present study—of combining semantic theory (namely semantic prosody) to the corpus-based methodology set forth by Baker et al. (2008) with Methodological Synergy.

Regardless of the 'weaknesses' highlighted in the discussion of the above two studies (Salama, 2011; Baker et al., 2008), they represent two very important steps in achieving a more even-handed and objective methodology for conducting CDA research. The research carried out by Baker et al. serves to lay the groundwork for the methodological change employed here by showing that CDA can effectively be complemented by corpus linguistics tools; while Salama (2011) took this methodological pairing a step further by

adding the use of collocations to the analysis process, thus—in theory—making the analysis more objective. However, where both studies seem to come up short is in relation to the selection of the data used. Even after beginning to integrate corpus linguistics into CDA studies, the data analysis has remained largely subjective and ideologically influenced. That is, although data collection approaches have begun to change, the data interpretation has not. Additionally, while Salama's research is more methodologically even-handed than Baker et al., it relies on the discourse found in one very specific area of language use (the books analyzed); at the same time, Baker et al. looked at a broader language area but examined the collocates of very specific, pre-determined LIs. Although different researchers have looked at these characteristics of the field in different ways, some of the most promising advances have come as a result of the use of semantic theory. Indeed, while both studies offer very important ideas for carrying out more objective forms of CDA research, they each have certain methodological weaknesses (even though they are small). However, in the current study the strengths of each approach are used along with approaches based in semantic theory in order to propose a better way to carry out similar analyses.

2.4. Application of a Semantic Approach

As can be seen in the preceding discussion, CDA has traditionally focused on language from an overtly subjective perspective. This is true both from the perspective of data selection and from that of data analysis. While studies which have utilized corpus technology have served as a response to this 'problem,' the issue of analysis remains an important one in being able to establish a truly objective form of carrying out CDA research. This being so, despite the fact that it has not been very frequently used in

research, there is little evidence to say that semantic theory cannot be applied to a CDA-influenced methodology, as has been shown by Freake, Gentil & Sheyholislami (2011), Salama (2011), Baker et al. (2008) and Orpin (2005). Using these studies as a basis, the present research intends to pair SP (the advantages of which are discussed below) with elements of Methodological Synergy (Baker et al., 2008). The combination of these two approaches will allow for CDA-style research to be carried out with minimal input from the researcher—thus responding to one of the principal weaknesses in CDA methodologies.

One of the most prominent instances of a fusion between Methodological Synergy and theoretical linguistics was a 2010 study by Prentice which looked at Scottish nationalist discourses using semantic tagging. This study looked at the language used by Scottish nationalist sources in portraying Scottish independence movements. Although normally this sort of study would be carried out with a simple textual analysis, Prentice utilized an electronic corpus to semantically tag certain linguistic features within the texts. This allowed Prentice to focus not on the interpretation of a few scattered instances of overt nationalist language, but on the general use of certain features (the use of *Scot*, for instance) throughout an enormous array of texts, while simultaneously examining how these features were presented in order to 'construct' a certain perspective or frame. This approach was taken by Prentice, "…on the grounds of [CDA researchers'] subjective application of a coding system that has been created to fit the data under investigation," (p. 408). Prentice goes on to suggest that,

> this criticism is only partially addressed by applying one's categorization system to a large volume of textual data, as there still exists a degree of subjectivity in the analyst's application of the categorization system. Automated semantic tagging,

> which offers us one possible way around this problem, and would allow relatively straightforward replication, involves the application of a computerized semantic coding system. (p. 408)

Semantic theory is one of the pillars of theoretical linguistics. Since semantic theory posits that LIs carry inherent meaning in any language, it follows that LIs are responsible for the communication of linguistic messages and, by proxy, the sorts of underlying discourses which are looked at in most CDA studies. Since semantics is generally seen as standing apart from context-based interpretations of language, semantic meaning is transparently present; its presence, then, dictates the underlying meaning present in a given use of language (either written or oral). To date, semantics has not been utilized extensively in studies dealing with CDA issues. Notwithstanding, there is a great deal of potential for its application in such areas. This is particularly true with regard to cross-linguistic studies.

Indeed, the present study is designed to experiment with and test this very application of semantic theory. Through the fusion of the approaches discussed in the preceding sections, it is proposed that a coherent, consistent, practical methodology can be created through which media discourse can be examined critically, but objectively.

One of the most obvious applications of semantic theory to CDA is through the use of semantic prosody. Generally speaking, semantic prosody in itself is what Stubbs (2007) calls, "…a relationship of habitual co-occurrence," (p. 1) or what Morley and Partington (2009) have described as 'evaluative harmony.' This is to say that—since semantic prosody is based on the analysis of collocations—it is predominantly focused on the idea that individual LIs have certain, consistent connotations (whether they be 'good,' 'bad' or

'neutral'). Further, these LIs and their respective collocates are seen to both have an effect on and be affected by the linguistic environment in which they occur (Zhang, 2010; Morley & Partington, 2007; Stubbs, 2001b). Although the general understanding of so-called evaluative harmony is somewhat divided insofar as scope (Hunston, 2007), there is general agreement in the field that LIs and their respective collocates both have an effect on and are affected by the linguistic environment in which they occur (Zhang, 2010; Morley & Partington, 2007; Stubbs, 2001b). In this respect, semantic prosody is well-suited to CDA research—especially in relation to the cyclical relation between discourse and reality which is discussed above (Carvalho, 2008) and in identifying frames.

Despite this potential for application, the studies which have been carried out using a semantic prosody-based methodology have been relatively limited. Since Louw coined the term 'semantic prosody' in 1993, the vast majority of studies which have been carried out using a methodology featuring semantic prosody have been dedicated to the analysis of single words or single phrases. While this is not the case in every instance of semantic prosody research (see, e.g., Ping-fang & Jing-chun, 2009; Zhang, 2009; Partington, 2007; Xiao & McEnery, 2006), there are relatively few studies which do not focus on semantic prosody according to use and semantic environment (Yusuf, 2010; Stubbs, 2005) or metaphor (Fillmore et al., 1988; Oster, 2010). Despite the fact that—at least at first glance—these types of semantic prosody studies are unrelated to the application being proposed here, they serve an important purpose in that they lay out the potential strategies, approaches, and forms of analysis that can be used in carrying out a study which utilizes a semantic prosody as a methodological tool.

At its most basic level, SP is the analysis of collocations, what Fillmore et al. (1988)—in reference to idioms—call, "familiar pieces familiarly arranged," (p. 510). That is, groups of words which consistently occur in groups within a given language. John Sinclair (1991)—one of the founding fathers of corpus and collocation research refers to two principles in collocation interpretation: the open-choice principle and the idiom principle. Both of these principles acknowledge that co-occurrence is a vital and universal feature of language; however, whereas the open-choice principle views language as a set of 'slots' waiting to be filled, the idiom principle conceptualizes is as a system in which certain combinations do not appear randomly (thus accounting for the idea of familiar arrangement put forth by Fillmore et al.).

To date, the most common usage of semantic prosody has been the analysis of idiomatic expressions and metaphors. While there have occasionally been innovative applications (Salama, 2011—in terms of the study's use of SP as a complement to CDA), analyzing the prosody of individual words within large corpora to determine their most general prosodies remains the most accepted application. However, whereas the analysis of metaphoric language focuses on seemingly unchangeable units of language, the examination of the collocates of individual words searches for more 'flexible' patterns within language. The analysis of single words as they occur throughout hundreds of contexts is one of the most frequently used forms of semantic prosody investigation. Indeed, it is this type of study which gave birth to 'Semantic Prosody' as a theory; it was actually Sinclair's realization in 1987 that LIs such as *happen* and *set in* almost always co-occurred with 'negative' events when looking at the Cobuild corpus which sparked the idea that semantic prosody might not only exist, but be systematically present within language

(see Zhang, 2010; Partington, 2004 for discussion). This, in turn led to Louw's famous 1993 study examining a few different LIs as well as coining the term 'semantic prosody' in reference to Firth's (1957) use of prosody to describe phonological characteristics which carried across words and sentences in speech giving the approach, "…its first definition, a 'consistent aura of meaning with which a form is imbued by its collocates'" (Louw, 1993 as cited in Zhang, 2010, p. 190). While Louw's study focused on the presence of irony in texts the field has since expanded to focus more prominently on questions of positive and negative connotations regarding habitual use of certain words in language.

These studies have confronted many uses of language, from looking at word uses as varied as the semantic prosody (SP) of the spoken use of *robot* (Yusuf, 2010) to things as simple as the use of the lexical item *fear* as a collocation used to portray emotions (Oster, 2010). However, regardless of the fact that there would seem to be a new crop of semantic prosody-related studies being published, the majority of readily available investigations continue to focus on broad, descriptive aims on the one hand (Stubbs, 2005; Partington, 2004) and research which shows more explicitly the presence of SP in language by using analyses focusing on irony (Partington, 2007) and cross-linguistic comparison on the other (Zhang, 2009; Xiao & McEnery, 2006).

Despite the fact that the field may seem broad in its approaches, it is actually quite limited in that the majority of studies simply focus on looking at semantic prosody and describing it within a given corpus. That is, there have been few practical applications using semantic prosody. Part of this is because the field is so young; however, this lack of application is also due to the sheer enormity of what can be looked at using SP. Nonetheless, some practical investigations have been carried out and these studies—along

with descriptive accounts of semantic prosody—serve to lay a solid foundation on which to base practical applications of semantic prosody. This is principally due to the fact that there has not been a solid proposal made regarding how to go about undertaking a semantic prosody study. This has, as a result, led to a general lack of methodological consistency within published studies.

Stubbs (2005), for example, carried out a study of Joseph Conrad's book *Heart of Darkness* using methodology very similar to what is alluded to in much of the published literature regarding semantic prosody. While Stubbs' research focused on themes very similar to those found in discussions of semantic prosody, the term itself does not appear once in the article. In fact, it does not even appear when discussing the contributions which John Sinclair made to the field. Nonetheless, the article is a fine example of the potential for application which semantic prosody possesses. Stubbs uses corpus linguistics to dissect the language use employed by Conrad, eventually culminating in a linguistically based discussion of the literary style employed in *Heart of Darkness*. Although the study does not explicitly address semantic prosody, many of the findings are reminiscent of semantic prosody-based research (i.e. Conrad's use of antonymous pairs and lexical contrast).

Although the Stubbs study (2005) is likely one of the most practical studies on semantic prosody-related themes, there are other semantic prosody-specific studies which feature interesting applications of the approach. Yusuf (2010), for example set out to find the prosody of the node word *robot*. Using corpus analysis tools, Yusuf found that in the vast majority of cases in which *robot* is found in spoken language it possesses positive prosody characteristics stemming from the fact that robots are used to aid humans and facilitate work. While there obviously are not a great many studies which specifically focus on

semantic prosody as a practical methodology for linguistics research, the studies which have been published do serve to establish a base on which to construct a new methodological approach to CDA using SP methodologies.

With this in mind, the present study combines some of the methodological tools discussed here (in relation to SP) and above (in relation to CDA and corpus linguistics). Through this synthesis, the study is able to focus on relatively small, stratified random corpora from a specific language area and carry out a CDA study of language use within these corpora using SP in a novel way. The methodology which was used here and the justification for it are discussed below.

3. Methodology

The following section presents an overview of the research steps which were taken in order to design, propose and test the new approach to CDA research which is discussed here. Generally speaking, this approach was made up of four hierarchically arranged, interconnecting parts. The steps which were undertaken were the product of a series of brief pilot studies (see Section 3.1.1, p. 58) and were designed to offer the researcher authentic data while simultaneously limiting the role that they would have in attaining and analyzing it.

First, two corpora were constructed. These corpora were each made up of texts selected to form representative samples of particular language communities' (American English and Mexican Spanish) print media discourse during a given month. In the present investigation, newspaper reporting was the only genre of print media examined. Second, a statistical analysis of lexical frequency was undertaken and a frequency limit was established. This was then followed by an examination of the collocates of frequently occurring LIs (node words); these were selected using the frequency 'ceiling'; and finally, those collocations which were found to be statistically salient were analyzed alone, as part of lexical groupings as well as both inter- and intra-linguistically. This was done through the use of a variety of means including frequency analysis and semantic prosody analysis within and across both corpora.

This multi-faceted approach was designed to analyze a stratified random sample of language use from a specific language domain. The set of steps used in the current study allowed for a methodology in which critical language was presented *to* the researcher

through its mere presence in the corpus data instead of one in which the researcher deliberately searched *for* language which was deemed (by the researcher) to be 'critical.' This characteristic of the present investigation is not only a hallmark of positivist research, but marks the most obvious deviation from the typical procedures utilized in many previous CDA-based studies (see, e.g., Orpin, 2005; Fairclough, 1995).

3.1. Overview of the Methodology Used

The current project utilized two principle methodological approaches (CADS and SP analysis) in order to explore the potential of an approach to CDA in which textual DA could be carried out as objectively as possible. Together, these approaches allowed for testing and carrying out a new approach to CDA research. Due to the methodological weaknesses which are laid out in the above sections with regard to both CDA-based studies and—to a lesser extent—SP studies, the current research project combined the use of CADS (Freake et al., 2011) viewed from a critical standpoint (e.g. Baker et al., 2008; Orpin, 2005) with semantic prosody analysis (Oster, 2010; Louw, 2008) in order to address the themes usually examined by CDA researchers from a new and more objective perspective. This combination of approaches was seen as complimentary for a number of reasons (discussed in detail in Chapter 2, p. 29) and was chosen in order to exploit the strengths of both approaches rather than focusing on their weaknesses.

Semantic prosody, for one, has very seldom been used for anything except the analysis of single lexical manifestations across mega corpora. As such, its use on a smaller, more practical level was seen as an important tool for effectively analyzing the findings gleaned from corpus analyses—something missing or under-represented in previous explorations of

CDA-corpus linguistics combinations (Salama, 2011; Baker et al., 2008). For example, while researchers like Oster (2005) have looked at varied occurrences of LIs related to *fear* or the concept of *corruption* (Orpin, 2005) using SP analysis, few, if any, have used SP as a tool to aid in the practical analysis of language and discourse.

At the same time, the use of corpus assisted CDA (and SP) on a more focused scale allows for a more complete analysis. This emphasis is something missing in many works, which tend to select a specific lexical grouping in a mega corpus (oftentimes a lexical item or synonyms for expressing a concept—similar to SP analyses) and analyze it in as many contexts as possible with the end goal of laying out the underlying discourses surrounding the unit analyzed (see, e.g., Salama, 2011; Baker et al., 2008). Although the analysis of a mega corpus can provide interesting findings, the data obtained from the corpus still has to be subjected to the researcher's subjective interpretation and analysis. Additionally, these sorts of studies lack the ability to effectively analyze particular genres and environments on a macro scale. This is partly due to features such as 'seasonal collocates' (Baker et al., 2008) in which certain language uses, "…are very frequent in a small number of years," (p. 286). With these methodological weaknesses in mind, the use of corpora based on stratified random samples from a specific area was complemented by the use of SP in that both methodological tools were made more practical and the resulting study was provided with a more complete theoretical foundation. Together, the two methods serve as a potential solution to many of the problems found in traditional approaches to CDA research.

Through the fusion of these two approaches, the present investigation looked to address two oft-cited weak points in CDA research: text selection and analysis criteria. As an

answer to the arbitrary selection of texts pointed out by Prentice (2010), common in many traditional CDA studies, texts were selected based on minimal topic-related lexical characteristics and were included in the final corpora only if they complied with certain criteria set forth at the beginning of the research process. Furthermore, in order to facilitate the analysis of large quantities of text, AntConc 3.2.4 (Anthony, 2011), a corpus analysis program was utilized in the examination of the final corpora of texts. As a response to the types of textual analyses often employed in CDA research, an approach based on SP analysis (particularly addressing the presence of collocations) was chosen in order to analyze the overall discourse present in the corpora. This not only grounds the corpus findings in theory but also highlights the general connotations of each corpus and, as a result, the media discourse in each country regarding the topic shared by both corpora. Moreover, steps were taken in order to make the approach to traditional CDA which is presented here as balanced and objective as possible.

To facilitate the methodological fusion used in the present study, various intermediary steps were taken in piloting, designing, assembling, collecting data from and analyzing the corpora. These steps, as they relate to different phases of the research process, are laid out briefly here and are discussed individually in more detail below. This study relied on the use and analysis of two corpora. These two corpora were assembled using topically parallel newspaper articles (focused on 'drug-related violence' in Mexico and the United States). All articles had been published during a single, randomly selected month during 2011 in newspapers from the United States (published in English) and Mexico (published in Spanish).

After having been constructed, the corpora were analyzed for lexical frequency, collocations and the Mutual Information (MI) scores of the collocates of salient node words. The use of MI scores in corpus linguistics gives the researcher a numerical representation of the 'strength' of a collocation and has been used previously in studies combining CDA and corpus linguistics as well as in semantic prosody studies (see Salama, 2011; Oster, 2010 for discussion). Mutual Information (MI) scores determine, "…whether there is a higher-than-random probability of the two items [the node being examined and a given collocate] occurring together," (Mautner, 2009, p. 125). This was done for all statistically salient LIs in both corpora.

After having created the final corpora, a raw frequency list was created for each corpus and was normalized to account for the different sizes of the two corpora. Upon establishing raw counts of lexical frequency for each corpus, a frequency 'ceiling' was established. Due to the fact that the corpus was only representative of a very specific area of language use and that there were a large number of LIs which only appeared once, the data was not evenly distributed. Thus, the frequency 'ceiling' was defined as a frequency of occurrence greater than the mean frequency of lexical occurrence. This allowed for analysis to be focused on only those LIs with a statistically salient presence in the data. After having established the frequency 'ceiling' for each corpus, a concordance analysis was carried out using LIs which were identified as 'frequently occurring' (this set of LIs excluded function words).

Having defined what constituted a frequent occurrence in the corpora, as well as which LIs had particularly 'strong' collocates, salient LIs were analyzed for their semantic prosody by observing their presence in the corpora based on their appearance as parts of

collocations and through the use of the concordance tool in the AntConc corpus analysis program (Anthony, 2011). This was done using the selected collocates' MI scores; leading the research to a point at which a general analysis of the semantic prosody of salient collocations could be undertaken. This analysis set the groundwork for the final goal of the study: to examine both corpora inter- and intra-linguistically based on lexical collocations and semantic prosody.

What follows is a detailed account of the methodology used in the present study. Because the current project is intended to be used as a starting point for future research, the methodological description is not laid out in a 'traditional' format. That is, the steps which comprise the method which was used are laid out chronologically as procedural steps. This was done for two principal reasons. First, the methodology which was used is a logical set of steps and to present it in a 'standard' format would prove quite confusing; when presented as logical steps, however, the method as a whole (as well as the rationale which informed its creation) can be more easily understood by the reader. Second, since the methodology which was employed here is intended to lay the groundwork for similar studies in the future, a step-by-step description of the methods employed not only allows for simple, straightforward replication as well as further testing, but for a transparent means through which to see any flaws within the method proposed here; thus allowing for the effective implementation of productive changes.

The experimental design which informed and was used in the present study is presented here, chronologically, in its complete form. The initial piloting process is discussed as well as its connection to the final research design. Following this, requirements for text inclusion in the corpora are detailed and their presence in and importance to the current

study is made clear. Additionally, the actual data collection and corpus construction is discussed in detail, particular attention is paid to the collection of texts, the corpora construction and the general characteristics of the corpora analyzed. Finally, a discussion of the statistical measures employed in analyzing the data is presented, as well as a description of the text analysis process.

3.1.1. Pilot Study

In order to be able to find and address the necessities for and methodological shortcomings of the current project, the first step which was taken upon having proposed research was to conduct a pilot study. Because searching for and selecting corpus materials was a vital step in establishing as much objectivity as possible, piloting was considered important as it provided an opportunity to experiment with different search methods while simultaneously becoming acquainted with the construction and analysis of text corpora. Being as the current project placed great importance on a degree of objectivity within data collection and analysis, it was of the utmost importance that the search terms (and not the *searcher*) returned representative and authentic data. This in turn allowed for the construction of representative corpora with little or no interference on the part of the researcher. The piloting period not only helped to gain experience at carrying out data collection, but also exposed many weaknesses in search methods which were later addressed when constructing and carrying out the final research.

The piloting period took place from October 28th through November 18th of 2011. During this period, one data collection was carried out per week. Pilot corpora were searched, constructed and analyzed four times during this period (Friday, October 28th;

Friday, November 4th; Friday, November 11th; and Friday, November 18th). Corpora on these days were constructed from texts published on the day in question, and were selected using *Google News* (discussed below) between 9 and 10 o'clock p.m. Texts included were found using set search terms (which evolved during the course of piloting—see Table 1) and had been published on the day of the search; they additionally were all written by credited authors. Once a search was carried out, all articles found were included in the pilot corpora. The corpora were then analyzed for raw frequency using the TextStat[5] corpus analysis program (Hüning, 2012) and the overall content of the corpora was examined. As can be seen in Tables 1 and 2, the use of search terms changed for both corpora during the piloting period. Initially, so-called 'wild cards' were used in order to have a broader set of results; however, this was eliminated from the final methodology; additionally, the final set of search terms was selected after experimenting with different manners of conducting searches during the first and second pilot runs. Because of this, the search terms were changed to those which were used for the remainder of the pilot study as well as in the final data gathering.

Table 1. Search terms used for English corpus piloting

Pilot Date	Search Terms Used	Corpus Size
Oct. 28	drug*/viol* (Advanced search)	2,646
Nov. 4	drug OR drugs AND violent OR violence	2,467
Nov. 11	drug OR drugs AND violent OR violence	4,228
Nov. 18	drug OR drugs AND violent OR violence	6,349

[5]The TextSTAT program was only used during the piloting process as the most important area of piloting was examining raw frequencies. Upon initiating the study itself, corpus analyses were carried out using AntConc 3.2.4 (Anthony, 2011). This program was selected because it offered a wider range of features—most notably, tools for statistical analysis such as MI scores; something not possible in TextSTAT.

Table 1 shows the search terms utilized in carrying out the pilot data collections for the English corpus. As can be seen in the table, the search terms evolved during the course of piloting and as they did so, the data analyzed was pulled from increasingly large corpora. Table 2 (below) presents the same information for the Spanish corpus piloting and demonstrates similar patterns of change in the search terms used and in corpora size.

Table 2. Search terms used for Spanish corpus piloting

Pilot Date	Search Terms Used	Corpus Size
Oct. 28	droga*/violen* (Advanced search)	3,303
Nov. 4	droga OR drogas AND violente OR violencia	2,234
Nov. 11	droga OR drogas AND violente OR violencia	2,041
Nov. 18	droga OR drogas AND violente OR violencia	8,133

In addition, the use of the Advanced Search setting in *Google News* was abandoned beginning with the third pilot run. This was deemed necessary because it was found that the use of a date range returned less reliable results (fewer results in general and many of them not from the date range being searched). Interestingly, this was only found to be the case in utilizing the Mexican version of *Google News*; there was no difference in the search terms when searching in English and so the Spanish terms were changed. The combined size of the pilot corpora used in the pilot study (texts from all data collection dates) was comparable to that of the final corpora which were used. The English pilot corpus was made up of 15,590 words (an average of 3,897.5 words per data collection[6]) and the Spanish corpus consisted of 15,711 words (an average of 3,927.75 words per data collection).

[6] *Data collection* refers to the set of data obtained from one search. That is, all the data compiled during each piloting date.

As is pointed out above, the main purpose of carrying out a pilot study for the current project was to fine-tune the search method used so as to distance the researcher from the data collection process as much as possible and to minimize direct involvement in changing, selecting, and assigning the data included in the final corpora. Nonetheless, it is impossible to ever completely remove a researcher from the research process. That is to say that a researcher still sets search terms, selects texts, decides on a topic for study and eliminates texts which are not representative of the discourse arena being examined.

With this in mind, the pilot portion of the study additionally served to highlight a few of the types of texts which were consistently found with the broad search terms used in the current study. So, although piloting helped to make the present study's final methodology more objective, it additionally highlighted certain points in data collection which would need to be addressed manually after data had been collected and corpora had been built. One of the most critical points highlighted through piloting was the need for individual revision of the texts selected using the initial search. Because one of the strong points of the present research is that, using the search terms established, very little had to be done in order to select texts, once the search terms were entered, texts related to the topic were found. However, through piloting it became apparent that, due to the minimal search terms used—the very feature which allowed for a degree of objectivity in text selection—it was necessary to manually double check to assure that the texts selected based on the search were representative of the corpora being analyzed. This was done by establishing a set of minimal characteristics for text inclusion during the pilot period.

In this sense, piloting was one of the most crucial parts of the entire research process. While certain characteristics of the present project are reminiscent of similar studies (see,

for example, Salama, 2011; Baker et al., 2008; Schrøder, 2007), both the care taken in order to assure as much objectivity as possible in collecting and assembling texts and the intention of being able to identify a discourse pattern in texts with no expectation of underlying ideologies would appear to be fairly uncommon in many CDA-based studies. As such, the pilot portion of data collection aided the present research by solidifying the manner of searching for, criteria for inclusion of, and criteria for exclusion of the texts which formed the final corpora used in the research.

3.1.1.1. Texts

Because the present research was based entirely on the simultaneous analysis of two separate, topically parallel corpora, the corpora were built using two sets of print media texts (in this case, newspaper articles) which were topically similar. The corpora which were analyzed were made up entirely of articles from major Mexican and American newspapers. The newspaper articles included in the final corpora met a specific set of characteristics which were representative of a particular topic: 'drug-related violence' in the United States and Mexico. The characteristics which informed the selection of texts as well as the lexical search terms (see Tables 1 and 2, pp. 59 & 60) served to eliminate any bias that would result if the corpora were arbitrarily constructed based on the topic being analyzed. The use of two linguistically distinct corpora covering parallel issues allowed for comparisons between the discourses used in each. This then permitted a comparison of each country's print media discourse (in the vein of Freake et al. (2011)); as opposed to a comparison between a specific discourse and language use in a mega corpus (i.e., Salama, 2011).

The possibility of choosing texts based solely on topic was acknowledged from the beginning of research as a potential strategy for corpus construction but was avoided as much as possible. This was done due to the influence which researchers have been shown to have on the outcome of a study of this sort if they exercise unilateral control over the texts chosen for analysis—especially when chosen based on topic (as shown in Poole, 2010). With this potential short-coming in mind—what Stubbs (1997) refers to as 'circularity'—it was deemed important to design the current study in such a way as to have the search terms used in data collection be 'responsible' for the brunt of data collection. Thusly, search terms were selected which reflected the underlying themes which informed the greater media discourse being studied here (e.g. Orpin, 2005). In this case, variations on drug- and violence-related lexical units were used for searching for and collecting data. Although these search terms evolved during the course of the pilot study (see Tables 1 and 2, pp. 59 & 60), they were consistently centered on the presence of LIs related to drugs and violence within print news articles, thus serving to almost completely restrict the texts which were collected to stories centered on the current 'drug-related violence' in Mexico and the United States. While the steps which allowed for this text selection were, are and will continue to be imperfect, the initial piloting process was immensely important because it allowed for the design of the data collection process which was used in the final research.

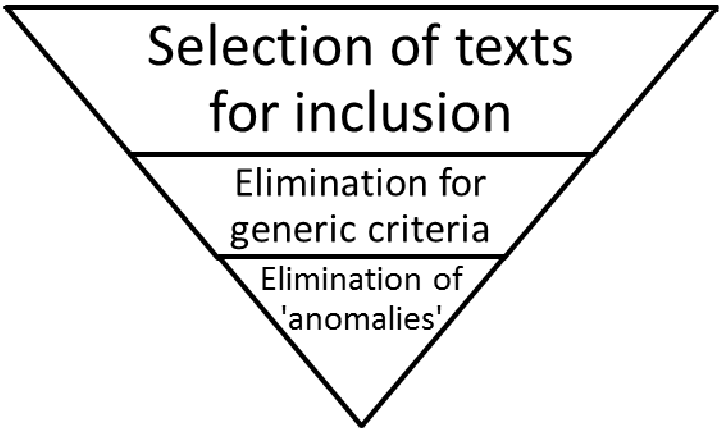
During the course of piloting it became apparent that there would need to be a further step employed in order to effectively limit the automatically collected data used in the final corpora. This was done only in the final data collection and was done in order to assure (based on findings from the pilot study) that the corpora were as representative of the language being studied as possible. Although some of the criteria for text selection were

established beforehand (the use of only ‘hard news’ stories, for instance) others were established based on methodological issues which were found during the course of the pilot study. The criteria used in compiling and constructing the final corpora are laid out in detail below.

3.1.1.2. Text Selection Criteria

In order to analyze the most representative corpora possible for the discourse use being studied, it was first important to establish the criteria which would be used to determine whether or not individual texts would or would not be included in the final corpora. This was done using categories which were established during the pilot study and which were applied after having first collected the data for the final corpora. That is, the search terms (established based on experimentation during the pilot study) were used to compile text corpora; following this, the criteria for text inclusion were applied and texts which fell outside of these criteria were excluded from the final corpora. In the interest of being redundant in selecting texts for the final analysis, the inclusion-exclusion process was carried out in three parts. Each of these parts was progressively less invasive and focused on more subjective criteria.

Figure 1. Process for determining inclusion in or exclusion from final corpora



The first part of the text selection process was the text search. This section of the process was the broadest and also the most objective part of the corpora building process in that it was dependent only on established search terms entered into *Google News*; further, all results which were brought up by this search were examined. Following the search portion of the research, texts were eliminated based on 'macro' criteria. The criteria which were employed in this step of the research process were intended to control for a variety of genre- and sample-based aspects of the final corpora. This was done in order to assure that the two corpora were as similar—and thus, comparable—as possible. While this section was more subjective than the previous step, since the criteria were based on the genre being examined (English- and Spanish-language hard news articles) and easily verifiable characteristics of the texts found in the first step, the researcher played a minimal role in actually 'deciding' on any inclusions or exclusions. In this sense, this step served as a sort of check list, articles were looked at individually and—based on the list of characteristics deemed necessary for corpus inclusion—were included or discarded based on whether or not they possessed the genre characteristics necessary for corpus inclusion (see Table 3, p. 67). Finally, after having constructed corpora which were relatively representative of the

generic and topical features being examined, a second round of eliminations was carried out before arriving at the final corpora (which were used for analysis). This round of eliminations (discussed below) was the most subjective step in the text selection process, but—once again—principally relied on the characteristics of the individual texts being looked at and not on the researcher's own opinions (at least to the extent to which such a thing is possible). Additionally, though more intrusive and subjective than previous steps, this step had the smallest impact on the overall makeup of the final corpora primarily because the majority of articles which could be eliminated already had been.

3.1.2. Sample Selection and Corpus Construction

In order to keep the corpora used in the current study as unbiased as possible, the most minimal criteria possible were utilized in establishing and building both corpora. This was done so as to avoid the very problems—characteristic of similar studies—which this project intended to confront (see Poole, 2010; Prentice, 2010 for discussion). However, these criteria were also established with care in order to ensure that the corpora remained representative of the topic being examined.

The articles which were examined were selected using *Google News* (see Section 3.1.1, p. 58). Through the use of a table of random numbers, a month in 2011 was chosen (October) and all articles included in both corpora had been published during that month. This was done because 2011 was the most recent full year. That is, all stories were as current as possible while still allowing for complete random selection in that articles could have been chosen from any month during 2011.

Upon having randomly selected October of 2011 as the sample month, articles were searched for and compiled using *Google News* by employing a set of parallel criteria for each corpus. Based on findings from the pilot stage, articles were searched day by day (that is, articles were not searched using a date range). This was done because during the pilot period a day by day search was found to yield more accurate and complete results than a search using a date range (i.e. an article search for publications between October 1st and October 31st). First, all articles which appeared on the *Google News* site for a given day were included in the initial sample. Following this, the list (all articles published in the month) was pared down using a set of generic criteria established for controlling the sample size and representativeness of the corpora.

Table 3. Criteria for inclusion in or exclusion from English and Spanish corpora

For Inclusion	For Exclusion
Published in Mexico or US	Published outside of US and Mexico
Published in October 2011	Non-newspaper (magazine, blog post, etc.)
Contain violen* and/or drug* (English)	Non-hard news (editorial, sports, human interest)
Contain violen* and/or droga* (Spanish)	Translations of stories from another country
Have a credited author	Published by wire service (AP, Reuters, 'redacción')

All stories which were included in the final corpora had to have been published online between October 1st and 31st of 2011. All stories included in the final corpora additionally had to have been written by credited writers. That is, no wire service stories, stories reprinted from wire services, or *redacciones* were included. The only exceptions to this stipulation were stories written by credited authors which were submitted to wire services and then distributed (i.e. a wire story with an author's name on the *by-line*). This was done in order to keep the corpora as representative of the discourse being studied as was

possible. Although wire stories obviously have writers, they go through a different editorial process and are written for a more general audience than are articles with credited reporters. Another exception was made for articles written by 'Staff' as this has begun to be a common practice in northern Mexico to preserve reporter anonymity due to violence perpetrated against journalists.

The only articles which were included in the construction of the final corpora were 'hard news' stories. These are stories which are commonly referred to as *headline* or *front page* articles. That is, no opinion, editorial, or human interest stories were included in the construction of the final corpora. Additionally, magazine articles, press releases, articles published by think-tanks, obituaries and stories published in countries outside of the United States and Mexico were excluded from the final corpora. This was done for reasons similar to those which informed the exclusion of wire stories from the current study.

Because the present research is being presented as a response to CDA research, it is important to focus on media discourse and the underlying social and political discourses which inform its development in the public eye. While opinion and human interest writing presents an accurate portrayal of current, popular discussions, they are too representative of 'colloquial' discussion to be included in the current research project. On the other hand, 'hard news' is—in theory, at least—a non-biased recounting of events. As such, these stories are more representative of the sort of discourse which CDA usually examines and, in isolation, represent a genre of print media in which (ideally) no underlying ideological discourse should exist. This is the case because of this 'representativeness,' as well as Carvalho's (2008) assertion that journalistic discourse intersects with every aspect of life

and that underlying discourse found in 'hard news' stories is important in that it will be taken at face value to be 'objective truth' by the audience being exposed to it.

In addition to these generic considerations, all articles used necessarily complied with the following micro-level lexical limitations which, in addition to allowing for an accurate and parallel method for searching and selecting articles, ensured that each corpus was comprised of articles reporting highly similar topics. This was done in order to control for the types of language and discourse topics being examined as much as possible. The first step in applying criteria to article selection was establishing search terms. Following a month-long period of weekly piloting (see Section 3.1.1, p. 58), during which time many approaches were taken to utilizing search terms within *Google News*, the following criteria were established for collecting Mexican and American newspaper articles.

- Mexican articles were searched using: *drogas OR droga AND violente OR violencia*
- American articles were searched using: *drugs OR drug AND violent OR violence*

The use of these particular search phrases was made necessary by the fact that Google does not allow one to search through the use of 'stemming.' That is, a search for *droga* will search every morphological variant of the lexical item (e.g. *droga*, *drogas*, and *drogada*). However, there is no way to search for variants of *droga* when it occurs as the stem of a given word. A search for *droga* will not yield results for a lexical item like *endrogada*, for example. Additionally, during the course of piloting the search terms to be used, it was found that searching *droga*—despite claims made by Google (2011)—limited results far more than if both *droga* and *drogas* were used. This was found to be the case when searching for *violente* and *violencia*, as well; this was the case when searching for

newspaper stories in English as well. The search for published articles containing these LIs formed the first part of the corpus construction process. The initial corpus size (all articles and headlines which appeared in a search—before delimiting them generically—see Table 3, p. 67) was 160 articles published in English and 160 articles published in Spanish.

Upon having selected articles from these initial corpora (based on the corpora selection steps laid out above), two separate corpora were constructed. The English corpus was made up of 33 articles with a total of 24,351 LIs. The Spanish corpus was made up of 24 articles with a total of 12,181 LIs.

Figure 2. Number of articles per day included in English and Spanish corpora

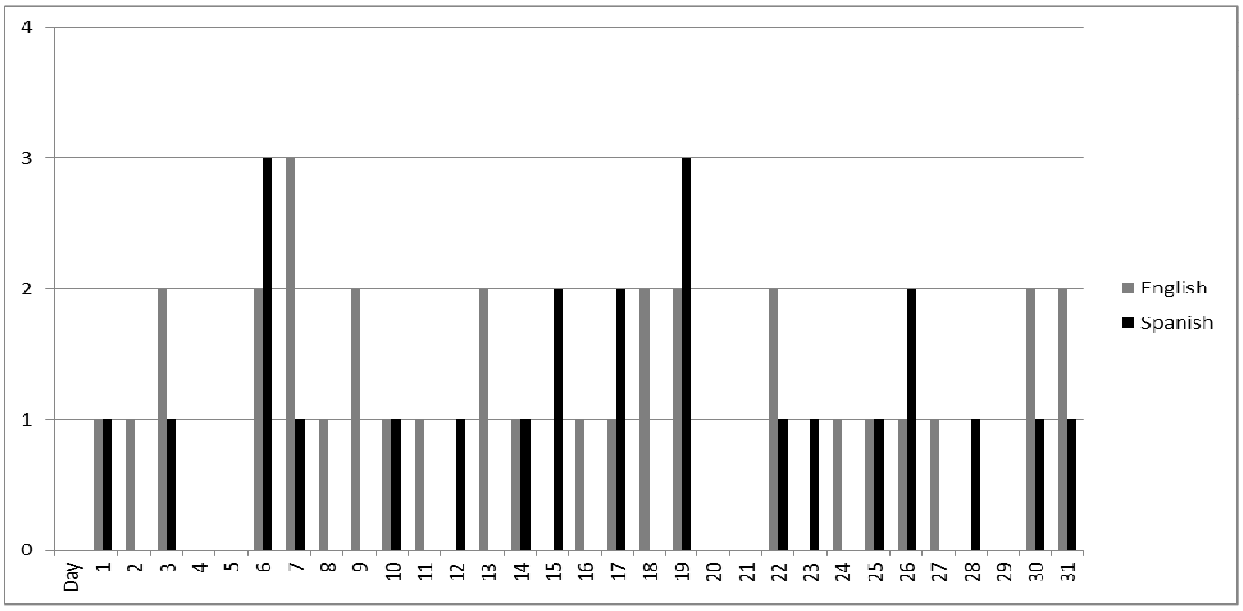


Figure 2 shows the number of articles used in the final corpora—according to the date of publication—before the elimination of 'anomalies' (discussed below). The articles featured in this phase of the corpus construction were those which had not been eliminated for generic features. In total, approximately 80.4% of English corpus articles and 85% of

Spanish articles were eliminated due to country of publication, genre, lack of a credited author, etc. (see Table 3, p. 67).

Based on the newspapers being searched as well as the search terms used, almost all of the texts encountered were related to current 'drug-related violence' in Mexico and the United States. However, there were some 'anomalies' within the texts collected which were eliminated prior to the construction of the final corpora. These 'anomalies' fell into two main categories which for the purposes of the current study were called *lexical* and *topical* anomalies. Although these would seem to be two separate categories of anomalies (and in some cases they were), most of the articles which were eliminated from the final corpora fell into both of these broad categories; lexical anomalies were articles which came up in the *Google News* search due to the presence of either *violen** or *drug** within the text, but not both. Though there were instances of lexical anomalies in which the article was still included in the final corpora (due to being on-topic), all of the texts which were eliminated for this reason were also topically anomalous (e.g. stories about prescription drugs, marijuana legalization or domestic violence). On the other hand, the texts which were eliminated for being topically anomalous were often only eliminated for topic. For example, two articles were eliminated from the final English corpus for being about a police operation to arrest members of a motorcycle gang. In this example, although the articles complied with the search terms established for inclusion in the corpora (both articles contained uses of *violen** and *drug**), topically the texts were not related to the media discourse being studied. As such, they were eliminated from the final corpora. All told, 11 articles (a total of 9,960 LIs) were eliminated from the final English corpus. No

articles were eliminated from the Spanish corpus, indicating that the news reported which featured one or both LIs had to do with the topic being studied.

Table 4. Lexical and topical elimination criteria for English and Spanish corpora

Lexical	Topical
drug*/droga* in non-illicit context	Occupy Wall Street "Bath Salts" Non-Mexican/American (same topic)
violen*/violen* in domestic abuse context	Afghanistan/Iraq Wars US-specific drug/crime stories

Table 4 shows the lexical and topical criteria for elimination from the final corpus. These were established during the pilot period based on the consistent appearance of certain topics. Some of the topical 'anomalies' were very closely related to the topic (e.g. US-specific stories), while others were found due to the lexical characteristics of the article(s) (e.g. those reporting on drug issues in other places). In the case of the latter three topical anomalies the topic being reported on was extremely similar to that being studied, but was eliminated (e.g. drug violence in Honduras). The first two topical anomalies, on the other hand, represented a form of what Baker et al. (2008) called 'seasonal collocates.' That is, topics which were briefly popular in the news media. In this case, both topical anomalies were found in the US corpus and were topics which were heavily reported during the fall months of 2011. Table 5 gives a brief overview of the actual articles eliminated from the final corpus due to being lexically or topically anomalous.

Table 5. Articles eliminated from final corpora

Number of Texts	Language	Topic
1	English	Soldier shooting
2	English	Arkansas drug arrests
2	English	Honduras crimes
1	English	Drug trial
2	English	Motorcycle gangs
1	English	Chinese boat traffic
1	English	Marijuana legalization
1	English	Occupy Wall Street
2	English	"Bath Salts"

Table 5 shows the articles eliminated from the final corpora according to language of publication and topic. These articles were eliminated due to both lexical and topical anomalies and included articles about the Occupy Wall Street movement, the use of 'Bath Salts' in the Northeastern United States, marijuana legalization and a Chinese military operation to stop the use of boats in the methamphetamine trade, among others. Interestingly, the American corpus was the only one which required articles be eliminated for these reasons. Although there were both topically and lexically anomalous texts in the Spanish corpus, they were all eliminated by the time this step was taken due to general criteria (see Table 3, p. 67) such as having been published outside of Mexico, or not having been written by a credited author.

3.1.3. Statistical Analysis of Corpora

Due to the nature of the present study, one of the most vital steps in carrying out a semantic prosody analysis of the text corpora was to first realize a statistical analysis of both corpora. This was deemed important to the current study in that a statistically based analysis serves as a response to many of the weaknesses present in past approaches to CDA

which have already been laid out and discussed. Beginning with the point of view put forth by Carvalho (2008), that an open reading (with limited pre-conceived notions of potential findings or goals of analysis) of texts "…allows for the identification of the most significant characteristics of the data…" (p. 166), it follows that a corpus analysis provides the researcher with a way to carry out this type of 'open' approach to analysis on a large scale. Further, the use of statistical analysis of corpora can permit the researcher to easily focus on LIs which occur with a statistically high frequency, thus allowing the research to more effectively examine the discourse present across texts by establishing 'pre-selection' (Oster, 2010). In this way, a consistent manner of text analysis could be established which was wholly dependent on lexical frequency and—by virtue—textual characteristics. Thus the researcher is able to study 'normal' occurrences within the text and not what they may have noticed or found interesting. This is an important step because it avoids the most common pitfall present in most CDA and semantic prosody studies in which the researcher sets out to analyze a particular word from the beginning of the investigation (thus, in one way or another, shaping the analysis itself) (see, e.g. Oster, 2010; Carvalho, 2008; Louw, 2008).

With these points in mind, the final corpora constructed for the present project were statistically analyzed for frequency and statistically salient LIs were then analyzed using the collocation search tool in the AntConc corpus analysis program (Anthony, 2011). This was done in order to focus on only those lexical features which were prominently featured in each corpus, thus avoiding the common CDA pitfall of "cherry-picking" which features to analyze (Mautner, 2009). Each corpus was subjected to two distinct statistical operations whose purpose was to bring lexical characteristics of each corpus to the surface. This was

done in order to permit pertinent data to 'appear' by virtue of its presence within the corpora. Statistical 'ceilings' were used for the initial frequency analysis and then for the collocation analysis (see section 3.1.4). Together, these 'ceilings' meant that the data examined could verifiably be seen as salient within the corpora.

3.1.4. Node and Collocation Selection

After having constructed both final text corpora, they were first analyzed for the raw frequency of each corpora's individual LIs. This was done using the AntConc corpus analysis program's 'Word List' feature (Anthony, 2011). Using AntConc, lexical frequency lists were obtained for each corpus using the 'treat all data as lowercase' setting in order to be as inclusive as possible in collecting data. In analyzing the corpora (particularly in regard to the frequency lists), the data was looked at in terms of 'types' and 'tokens.' This distinction, according to Kennedy (1998), is one primarily based on a word's underlying function within a corpus; 'types' being individual items and 'tokens' being occurrences of individual types. For example, two distinct morphological realizations of a single LI (e.g. *drug* and *drugs*) are seen to be two tokens of a single type. Nonetheless, for the purposes of the current study tokens were almost exclusively used in analysis. This was done because the broad presence of LIs was what was deemed most important and not merely whether they were present or not. For the analysis of American print media, the corpus used was made up of 14,391 tokens. The Spanish language corpus was comprised of 11,982 tokens.

A frequency list was made first in that it allowed for the remainder of the research steps to be taken effectively. After the initial frequency list was made, all single occurrences were eliminated from it since a single occurrence could not really offer

anything to a study of frequent occurrences. That is, while a LI which appeared once might serve an important purpose within a discourse, without multiple occurrences there was no way to extrapolate conclusions from its presence (for the purposes of the present methodology). Specifically, an item with a frequency of one could be seen as a random occurrence. LIs with a frequency of one were by far the most common features within both corpora. In the entire English corpus, 55.5% of LIs only occurred once; while in the entire Spanish corpus single occurrences accounted for 59.8% of the total number of frequent occurrences.

Table 6. Distribution of lexical items according to frequency in the English corpus

Freq. of Occurrence	Freq. per 1,000 words	% of Total
1 through 3	.0695-0.2085	79.87%
4 through 6	0.2770-0.4169	9.70%
7 through 9	0.4864-0.6524	3.40%
10 through 14	0.6949-0.9728	2.69%
15 through 20	1.0423-1.3898	1.42%
21 through 50	1.4593-3.4744	1.75%
51 through 1,000	3.6134-63.9288	1.13%

Table 6 shows the distribution of LI frequencies in the English corpus. As can be seen, the vast majority of LIs occurred between one and three times in the entire corpus. Table 7 (below) shows the same information for the Spanish corpus.

Table 7. Distribution of lexical items according to frequency in the Spanish corpus

Freq. of Occurrence	Freq. per 1,000 words	% of Total
1 through 3	0.0835-0.2504	85.50%
4 through 6	0.3338-0.5008	7.88%
7 through 9	0.5842-0.7511	2.61%
10 through 14	0.8346-1.1684	1.64%
15 through 20	1.2519-1.6692	0.57%
21 through 50	1.7526-3.6722	0.78%
51 through 1,000	4.7571-78.701	1.77%

After having eliminated all single occurrences from the corpora, the general lexical frequency of each corpus was determined in order to establish a frequency 'ceiling.' This was deemed to be the average frequency of lexical occurrence of the remaining tokens for each corpus (9.25 for the English corpus and 8.53 for the Spanish corpus). As such, the LIs included in the present study occurred with a frequency of ten or higher in the English corpus and nine or higher in the Spanish corpus (with a minimum of 0.6524 occurrences per 1,000 words in English, and 0.7511 in Spanish); having done this, the corpora used for final analysis were considered to be representative of the most statistically salient LIs in each language.

After finding the most statistically salient LIs for each corpus (115 LIs in English and 71 in Spanish), all function words and single letter tokens were eliminated from both corpora (Salama, 2011; Orpin, 2005), thus allowing for more focused analysis in that these LIs were, by and large, the most frequently occurring in each corpus. Both of these items were eliminated for similar reasons, namely that they served a syntactic purpose in the corpora but not a semantic one. The elimination of function words is standard practice in CDA research because they are devoid of meaning and thus serve little purpose in meaning-

based analyses (Mautner, 2009). Similarly, the majority of single letter tokens were grammatical (i.e. plural and possessive markers or parts of contractions).

Additionally, the elimination of these features helped to make both corpora vastly more manageable for analysis. For instance, the article *the*, in the frequency list for the English corpus, occurred nearly nine times as frequently (920 appearances) as did the most frequently occurring content word, *drug* (108 times). Similar characteristics were found in the Spanish corpus where the preposition *de* occurred more than ten times as frequently (943 times) as did the most common non-function word *México* (86 times). Due to the nature of the study being discussed here, both of these steps were included in order to make the total amount of data to be analyzed more manageable and also more lexically transparent. That is, the data which was analyzed could be said to (a) be representative of the most statistically common lexical uses in each corpus (through the use of a frequency 'ceiling') and (b) be representative of meaningful discourse (through the elimination of function words).

The initial frequency lists for the English and Spanish corpora contained 3,081 and 2,981 LIs, respectively. After having eliminated both single token occurrences and function words, the English frequency list contained 115 salient LIs and the Spanish list contained 71. The lists which were obtained at this point served as the corpora from which to draw salient nodes (see below) for the collocation and SP portion of the analysis. A list of the node words used can be found in Appendix A (see p. 135).

3.1.4.1. Sorting of Nodes and Collocations

Once the salient node words were selected from each corpus, the final phases of analysis could begin. This process involved the use of the AntConc corpus analysis program (Anthony, 2011) to search concordances for all node words, the establishment of a 'ceiling' for determining significance of collocate strength and the actual comparison and analysis of the language used in both corpora. The following sections present the methodological tools used as well as a discussion of their importance to and use in the present study.

3.1.4.1.1. Node Words

In all studies which examine the use of certain LIs as part of a KWIC search (such as the present study) the basest unit used in the analysis is the 'node word' (Baker et al., 2008; Xiao & McEnery, 2006; Sinclair, 1991), sometimes also called the 'key word' (see Orpin, 2005). The node word is a given LI which is studied alongside its collocates (Sinclair, 1991).

Although there are multiple ways to refer to a LI within a corpus as well as its classification within said corpus, the most widely used distinction is that which separates 'types' from 'tokens' (Kennedy, 1998). In the present study, the AntConc corpus program (Anthony, 2011) was used to analyze the presence of both types (all lexical features in the corpora) and tokens (those present in the frequency list). While this distinction is an important one in laying out the groundwork for a lexically based corpus analysis (such as in Kennedy's case), for the purposes of the current project, the brunt of the methodological focus was placed on what Kennedy (p. 251) calls the, "…target item, node word or search

item." Although Kennedy and many other authors have alternately used the term keyword as a synonym (Prentice, 2010), the terms 'node' and 'node word' will be employed here as they are most commonly used in studies which principally focus on concordance analysis (see Salama, 2011; Oster, 2010; Louw, 2008).

The node words which were found through the previous methodological steps (Section 3.1.4, p. 75) were used as the basis for collocation analysis. That is, they served as the nodes for the KWIC portion of the present study (see Tables 8 and 9). Tokens which are being examined in their capacity as node words are included in all capital letters.

Table 8. Top ten most frequently occurring nodes in the English corpus

Node Word	Raw Freq.	Freq. per 1,000 words
DRUG	108	7.50
MEXICO	99	6.88
BORDER	66	4.59
MEXICAN	61	4.24
CARTEL	60	4.17
STATE	52	3.61
OFFICIALS	45	3.13
POLICE	44	3.06
ONE	40	2.78
VERACRUZ	39	2.71

Table 9. Top ten most frequently occurring nodes in the Spanish corpus

Node Word	Raw Freq.	Freq. per 1,000 words
MÉXICO	86	7.18
VIOLENCIA	59	4.92
DROGAS	57	4.76
ESTADOS	44	3.67
PAÍS	39	3.25
UNIDOS	33	2.75
SECUESTRO	32	2.67
GRUPOS	28	2.34
AÑOS	27	2.25
ESTADO	27	2.25

Tables 8 and 9 show the ten most frequently occurring nodes in each corpus along with the raw frequency and a normalized frequency of occurrence for each of them; all nodes included here (as discussed above) were content words which occurred at an above average rate of frequency in each corpus (see Section 3.1.4, p. 75). Upon having drawn up a list of all the nodes to be used, the next step in the research process was a concordance analysis of these.

3.1.4.1.2. Concordance Analysis

Having found pertinent node words for each corpus, these were then analyzed for concordance features using the AntConc corpus analysis program (Anthony, 2011). This was a necessary step in carrying out a SP analysis of the data contained in both corpora (see Louw, 2008; Partington, 2004). As Partington points out, the cornerstone of SP theory hinges on the idea that most words have inherent evaluative meanings (i.e. negative, positive, or neutral connotations) and that these can and are used with specific ends depending on the author or speaker's intentions. Because of this, it is possible to examine the consistent in-text behavior of certain LIs using corpus analysis tools in order to observe the prosodic characteristics of a given word. One of the most frequently used ways of going about this is to examine the collocates of specific LIs to analyze patterns in their use. This has been done frequently by many authors (a thorough discussion of this can be found in Partington (2004)). The present study employed a concordance analysis using salient node words—the selection of which is described above.

Because semantic prosodies are formed based on the co-occurrence of certain LIs (Louw, 2008), it was first important to determine what some of the salient co-occurrence patterns were in the corpora used here. With this in mind, every node word in each corpus (135 English LIs and 65 Spanish LIs) was searched using the *Collocates* tool featured in the AntConc corpus analysis program (Anthony, 2011). This feature allowed for nodes to be searched within a given corpus and presented the collocates of each. The tool also can be adjusted to accommodate different spans of text around a node, and can show how frequently the node and collocate occurred together and how often a given collocate occurs before or after the node; the tool also features a statistical tool which can be used to analyze different facets of a given search term's co-occurrence patterns.

For the current research, the concordance analysis was comprised of two main methodological processes. As with previous research steps, these two points focused on separating the most salient data possible while simultaneously relying on as little subjective input from the researcher as possible. The node words were searched according to some of the stipulations set forth by Salama (2011, citing Hunston, 2002). First of all, collocates were searched for within a five-word window on either side of the node words. For example, in a search of the node MEXICO in the English corpus, the tokens which appear on either side of the node in the sample KWIC analysis presented below (Figure 3) represent the environment within which collocations were searched by the program.

Figure 3. Sample KWIC presentation from English corpus (MEXICO)

Beltran Leyva cartel, according to	**MEXICO**	City s Attorney General Miguel Ángel
dividing the city from neighboring	**MEXICO**	state along a busy road by the Defense
yet to find the bodies. While	**MEXICO**	's on-going drug war has made its way

As can be seen in Figure 3, the corpus analysis program shows the node word in context. Having done this it is then possible to examine the collocates which frequently co-occur with the node. In the case of Figure 3, one can easily see that the collocate *city*, for example, co-occurs twice with the node in three contexts, suggesting a strong collocation. Based on this type of analysis, once a search is entered, the AntConc program (Anthony, 2011) produces a list of frequent collocates along with all pertinent information about them, their presence in the corpus, and how they co-occur with the node word being searched.

The present study was carried out using a list of statistically determined node words from each corpus. These words were searched for one at a time using the AntConc corpus analysis program (Anthony, 2011). In order to search the node words effectively and in a parallel manner within each corpus, the following steps were taken and the settings employed are described. In the interest of including all possible tokens of a given type, LIs were searched using the 'treat all data as lowercase' setting in AntConc (Anthony, 2011). This means that a search for the node CITY would generate results for the LI's use as part of a name (as in, *Mexico City*) as well as a noun used to describe a populous grouping of inhabitants bigger than a 'town' (both instances of this node's use are present in the above KWIC—see Figure 3, p. 82). This insured that no use of a given LI would go unnoticed when working with the corpora. Additionally, the data shown and consequently used in the concordance analysis had to have been present two times or more within the corpora. This decision was very similar to the choice to eliminate single frequencies from the frequency list and was used in the data collection portion of the research for the same reasons. Namely that no matter how strong a given collocation is found to be, it becomes difficult to infer anything about its use should it only occur once in a corpus.

In order to analyze the strength of the individual collocations found through these steps, MI scores were used. MI scores are a form of 'association measures' (Bouma, 2009). The use of MI scores allows the researcher, "…to rank candidates extracted from a corpus…" (p. 58). Upon having selected the top ranking 'candidates,' these can be seen as co-occurring significantly and can then be analyzed. The use of MI scores has been used in many similar corpus analysis projects (see Salama, 2011; Oster, 2010; Xiao & McEnery, 2006) and has been found to be quite effective in highlighting patterns of co-occurrence in many differently sized corpora (Xiao & McEnery, 2006). The use of MI scores provides the researcher with a number corresponding to the 'strength' of a given collocation. In cases of MI's use in similar studies (see Salama, 2011; Xiao & McEnery, 2006) the number considered representative of significant collocation has been three or higher (indicating a co-occurrence which can be said to have not occurred by chance). In the interest of limiting the size of data output and being as methodologically rigorous as possible, the current study examined those collocates found to have an MI score of four or higher. Thus, all collocates analyzed here had both a MI score of four or higher and co-occurred with their corresponding node two or more times within the corpus. Upon having extracted only these collocates from the full list of collocates for each node, the actual analysis of each corpus' language use could begin.

3.1.5. Final Analysis of Corpora

After having constructed and statistically analyzed both corpora and having searched and sorted the collocates of all statistically salient node words found therein, the final step in conducting the analysis was to look at the data obtained from a SP perspective. This was done in two phases. First, data was looked at according to the raw frequency of certain LIs

(analyzed as nodes); and second, any node words which occurred in both corpora were examined side by side. Although when looked at as part of the entire analysis process (along with corpus construction, and both frequency and collocation analyses) the SP-based portion of the present study may seem to make up a small part of the methodology, it was the culmination of the preceding methodological steps.

In the present project, the goal was to locate instances of SP in the corpora themselves. This was accomplished by examining the data collected in both corpora using the AntConc corpus analysis program (Anthony, 2011). The first part of the SP analysis involved the examination of the most frequently occurring LIs and their collocates in both corpora. Following this, the same analysis was carried out using any LIs which had an equivalent in the other corpus (e.g. DRUG and DROGA or MEXICO and MÉXICO). SP characteristics were determined for each item based on linguistic competence. That is, since prosodic characteristics are natural in all words, competent speakers are able to determine these features based on experience. While this has been debated in SP circles (see Hunston, 2007), there is a wealth of information to substantiate the existence and consistency of SP's presence in language. SP can be seen in many simple examples of language use.

Obviously, a competent speaker of English will never say that a word like *happiness* has negative connotations and, likewise, a competent Spanish speaker will not say that a word like *muerte* has positive connotations unless speaking out of context or employing humor. However, this can be seen even clearer in looking at the use of linguistic irony. SP is particularly obvious in irony precisely because many uses of linguistic irony are accomplished by invoking the opposite prosody in a word (e.g. sarcasm) (see Partington (2007) for complete discussion).

The analysis of the SP characteristics of node words and their collocates was by far the most novel part of the present study in that not only has SP not been extensively studied, but it has rarely—if ever—been applied to CDA and corpora in the way that it was here. Essentially, once a set of collocates was extrapolated for a given node word, they were examined for their prosodic features, both on their own and in relation to the linguistic environment in which they were found to be present in the corpora. As in other studies utilizing SP as a methodological tool, prosodic characteristics were seen as part of a given word's "DNA" (Morley & Partington, 2009) that is, there is no discussion as to whether a word is positive or negative; it simply is. Being as there is no way—at least presently—for a researcher to determine a LI's prosodic characteristics without involving themselves in making a determination, that is precisely what was done here. As a competent speaker of both Spanish and English, the prosodic characteristics of individual LIs were determined by the researcher. Words known to have a positive prosody were deemed as such and words deemed to be 'negative' were treated likewise. Any words of which there was doubt were treated as neutral (these were principally titles and geographical terms). This methodological process, as well as the results which were obtained using it, is discussed in more detail in the following chapter.

4. Results

In the following section, the results of the present study are presented and discussed. The data analyzed came from two corpora constructed from the newspaper discourse present in two countries regarding the same topic ('drug-related violence') during October of 2011. Because the main ambition of the current study was to explore the feasibility of such research, the results presented here are only a selection of pertinent findings pulled from all of the data gathered.

The results are divided into two sections. These sections approach the corpus data on two levels. The first level of analysis focused on investigating the predominate discourse in both corpora, while the second concentrated on the linguistic similarities and differences encountered in examining the shared concepts found in both corpora. The corpus data were examined and presented in this way in hopes of effectively addressing certain weaknesses within CDA-based research. This direction was taken because it was these very methodological weaknesses which served as the basis for designing and conducting the present study.

The first section of the results discusses findings related to the Semantic Prosody analysis of the most frequently occurring node words in each corpus. This was done in order to respond to the accusations which are quite commonly leveled against CDA (see Poole, 2010; Mautner, 2009) that researchers are often liberal in selecting the data to be analyzed in CDA investigations. Through a focus on the most commonly used linguistic types in the corpora, the predominant discourses in each corpus could be analyzed based

only on salient lexical features and not on preselected search terms (Salama, 2011; Baker et al., 2008).

The second section discusses the prosodic characteristics found while examining node words which were present in both corpora (these were called, for the purpose of this study, 'shared concepts'). This analysis was included in order to explore the similarities and differences in the presentation of certain types within both corpora. Additionally, this approach helped to test not only the possibility of carrying out a study using the methodology employed here, but also helped to determine the applicability of this sort of methodology in analyzing multi-lingual corpora.

Together, these two approaches to analysis provided a balanced perspective through which to examine the general characteristics of each corpus as well as to compare and contrast the discourse characteristics present in both. As a result of this approach, it was possible to analyze the general print media discourses present in both countries from inter- and intra-linguistic points of view. The findings obtained through these two approaches are presented in the following two sections.

4.1. Frequent Node Words

The examination of the most frequent lexical items (LIs) in each corpus was one of the least labor-intensive portions of the data interpretation as a whole. Despite this, it was also one of the most important parts of the entire data analysis process. This is because the analysis of frequent node words not only showed the presence of salient LIs within the discourse of each country's print media, but also selected the nodes to be analyzed (as a response to claims of selective data analysis in CDA) (see Poole, 2010; Mautner, 2009).

Because of this, the analysis here was expansive in its application of SP as an analysis tool. That is, broad discourse trends in the corpora were examined more than were minute individual differences or similarities (see Section 4.2 for discussion of these, p. 101). This was done because the most frequent LIs in each corpus could be seen to be representative of each country's print media discourse regarding 'drug-related violence.' Because of this, an overview of the prosodic characteristics of the most common nodes gave a balanced overall impression of the discourses found in each corpus.

For this section, the top ten most frequently occurring LIs for each corpus were selected and their collocates were analyzed. Table 10 shows the top ten most frequent LIs for both corpora along with the actual lexical frequency and the normalized frequency for each (normed frequency per 1,000 words according to corpus size).

Table 10. Top ten most frequent LIs in both corpora

	Node	Freq.	Freq./1,000 words		Node	Freq.	Freq./1,000 words
English Corpus	DRUG	101	7.02	**Spanish Corpus**	MÉXICO	86	7.18
	MEXICO	98	6.81		VIOLENCIA	58	4.84
	MEXICAN	61	4.24		DROGAS	55	4.59
	CARTEL	57	3.96		PAÍS	37	3.09
	BORDER	55	3.82		ESTADOS	33	2.75
	STATE	48	3.34		UNIDOS	33	2.75
	POLICE	43	2.99		SECUESTRO	29	2.42
	OFFICIALS	42	2.92		GRUPOS	28	2.34
	VERACRUZ	39	2.71		AÑOS	27	2.25
	VIOLENCE	38	2.64		MEXICANOS	25	2.09

As can be seen in Table 10, nearly all of the most frequent LIs from each corpus were directly related to the topic of 'drug-related violence.' Even those LIs not obviously related to the topic (e.g. *años* or *state*) were found to be tied to the topic through the context in

which they were found to be used in the corpora. This was seen as a demonstration that the corpora constructed were in fact representative of the language-use being examined. This can also clearly be seen in that some of the search terms which were used in collecting data for the corpora were also found to be present in the top ten most frequent LIs for each corpus.

The first, most general, analysis of the data was carried out using each corpus' top ten most frequently occurring node words. This portion of the study involved nothing more than an overview of the connotative characteristics that each corpus had, based on the understanding that individual words carry positive, negative or neutral connotations in their 'DNA' (see Morley & Partington, 2009). In order to examine the prosodic characteristics of each corpus, the prosodic characteristics of the most frequent LIs (see Table 10, p. 89) and their collocates were examined. In order to examine the broad prosodic characteristics of each collocate in relation to its node and the discourse itself, the 15 most frequently occurring collocates were selected for each node (as in the node selection, function words were not included in the analysis) and these collocates were analyzed according to their individual prosodic characteristics as well as their presence within the corpora (using concordance analysis). This is discussed below.

As was expected, the most frequent LIs in both countries' corpora were thematically related to the topic of 'drug-related violence.' However, it is important to note that in only examining the raw frequency of overtly positive, negative or neutral connotations of the nodes presented here, both corpora appeared to give relatively equal representation to the same nodes. Also, the corpora appeared to be balanced in terms of the use of certain prosodic characteristics. The US corpus contained three negative nodes, DRUG, CARTEL

and VIOLENCE. Similarly, the Mexican corpus contained VIOLENCIA, DROGAS and SECUESTRO. Though any difference in this respect was slight, it is interesting to note that the only difference between both of these sets of prosodically negative nodes was the US's use of CARTEL and Mexico's use of SECUESTRO.

As was pointed out above, the presence of VIOLENCE/VIOLENCIA and DRUG/DROGAS is not in itself noteworthy in that these were search terms used to compile the data used in the corpora. However, the difference in the other negative node in each corpus is interesting since, in theory, both corpora are made up of texts regarding the same topics (VIOLENCE and DRUGS). These frequent LIs function like a window into the predominate discourse off each corpus (see above). Since both corpora share the common theme of 'drug-related violence' it is not surprising that these concepts are well represented in looking at frequent nodes. As such, there would seem to be a split between the overriding discourse of each country's print media in that there is a difference of representation regarding these nodes. This was seen in other frequent nodes as well (see the discussion of BORDER/FRONTERA, p. 95). A few noteworthy examples of these sorts of similarities and differences are discussed below, particularly the collocation with the token *security* (see p. 96).

In both corpora the most frequently occurring node words were, by and large, neutral (many being place names). Nonetheless, most of these nodes frequently co-occurred with negative collocates. MEXICO/MÉXICO, for instance is a place name and, thus, can carry neither overtly positive nor negative connotations on the part of the author(s) who printed it. Despite this, the nodes MEXICO/MÉXICO were found to co-occur with a number of negative collocates in both corpora. Table 11 shows the 15 most frequently occurring collocates of the node MEXICO in the English corpus along with the MI scores (see

Section 3.1.4.1.2, p. 82) and the raw frequency and joint frequency[7] for each. Table 12 (below) shows the same information for the node MÉXICO from the Spanish corpus.

Table 11. Collocates of MEXICO in the English corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
MEXICO	drug	108	14	4.24
	city	37	13	5.67
	violence	38	9	5.11
	northern	11	7	6.53
	cartel	60	7	4.08
	violent	15	5	5.60
	veracruz	39	5	4.22
	two	34	5	4.42
	states	35	5	4.38
	texas	29	4	4.33
	several	14	4	5.38
	american	23	4	4.66
	years	19	3	4.52
	war	15	3	4.86
	troops	10	3	5.45

[7] *Raw frequency* and *joint frequency* refer to the manner in which a given collocate is present in relation to the corpus as a whole as well as to the node with which it collocates. *Raw frequency* refers to the number of times that a LI appears in the entire corpus, while *joint frequency*, specifically refers to appearances as a collocate of the node being examined.

Table 12. Collocates of MÉXICO in the Spanish corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
MÉXICO	drogas	57	10	4.61
	violencia	59	8	4.24
	unidos	33	4	4.08
	tráfico	17	4	5.03
	nuevo	10	4	5.80
	grupos	28	4	4.31
	mayor	14	3	4.90
	líder	6	3	6.12
	cárteles	22	3	4.25
	contra	21	3	4.31
	ciudad	18	3	4.54
	unido	2	2	7.12
	texas	4	2	6.12
	tema	10	2	4.80
	sucede	2	2	7.12

As can be seen from looking at the collocates for each node, many of the LIs which co-occurred with MEXICO and MÉXICO had negative connotations. In solely looking at raw frequencies, the American corpus featured a much higher number of negative collocates than did the Mexican corpus. Nevertheless, frequently occurring collocates did not necessarily have high levels of joint occurrence. For example, a negative collocate such as *violencia* is, statistically, a collocate with the node word MÉXICO; even so, *violencia* only occurs as a collocate (within five tokens on either side) of MÉXICO eight times despite being present 59 times in the entire corpus.

In the interest of maintaining as balanced of an analysis as possible it was also important to view overtly positive items which co-occurred with the nodes being examined. In the case of the node word MÉXICO, two of the collocates which featured high levels of joint occurrence with the node word MÉXICO also had positive prosodic characteristics

(*líder* and *unido*). Despite this, both occurred in negative syntactic contexts within the Mexican corpus.

In the case of *líder*, the collocate co-occurred with the node MÉXICO in reference to either leaders of drug cartels or Mexico's place as the global leader in drug trafficking. Similarly, the LI *unido*—despite a positive prosody—was only featured (in the entire corpus) as part of the name of an organization, *México unido contra la delincuencia*; showing that even as a positive term, within the corpus being examined here, *unido* is a strong collocate of the negative LI *delincuencia*[8]. These collocates are important to the analysis of both corpora because they exhibit some interesting behaviors. What is remarkable about them is that they are present to begin with (the US corpus featured no positive collocates for MEXICO—see Table 11, p. 92); but it also is notable that even though they appear, they did not actually make the discourse surrounding the node MÉXICO any more positive (at least according to the analyses carried out here).

The frequency of individual nodes within each corpus was additionally important in that (aside from SP analyses) their occurrences alone offered insight into the general discourse of each corpus. This can be clearly seen in the case of the node word BORDER. BORDER was one of the top ten most frequent nodes in the English corpus. However, this was not the case for its Spanish equivalent FRONTERA. In fact, when looking at the normalized frequencies of each node word, BORDER occurred over four times more frequently than did FRONTERA (4.6 and 1 occurrence per 1,000 words, respectively). While this is a key indicator of how important the concept of the border between Mexico and the US was in

[8] In some respects, *unido* can be seen—in this example—as presenting positive prosodic characteristics. While debating this point is beyond the scope of the current project, it bears pointing out that even though *unido* is representative of a generally positive concept, the fact that it co-occurs as a sort of opposition to crime implies that it is negative within the data analyzed here.

the print media discourse examined here, it is also interesting to note the collocates which were found to frequently co-occur with the nodes BORDER and FRONTERA. Table 13 shows the significant collocates of the node BORDER in the English corpus along with the raw frequency, joint frequency and MI score for each.

Table 13. Collocates of BORDER in the English corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
BORDER	mexican	61	9	5.01
	security	30	8	5.86
	mexico	99	7	3.95
	patrol	8	6	7.35
	drug	108	5	3.34
	agent	12	5	6.51
	years	19	4	5.52
	one	40	4	4.45
	long	9	4	6.60
	governor	8	4	6.77
	threatening	4	3	7.35
	texas	29	3	4.50
	state	52	3	3.65
	shelters	3	3	7.77
	perry	33	3	4.31

Table 14 shows the significant collocates of the node FRONTERA in the Spanish corpus along with the raw frequency, joint frequency and MI score of each.

Table 14. Collocates of FRONTERA in the Spanish corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
FRONTERA	unidos	33	3	6.50419
	norte	9	3	8.37866
	estados	44	3	6.08915
	méxico	86	2	4.53735
	lado	7	2	8.15626
	ambos	4	2	8.96362

At first glance, the most conspicuous difference in the collocates found for each node is that the Mexican corpus did not even have 15 frequent collocates. This is partly the case because BORDER was so much more common in American media discourse than it was in Mexican media discourse. Still, it is important to note that the LIs which collocated with BORDER tended to be related to the discourse of 'drug-related violence,' whereas the node FRONTERA principally co-occurred with collocates that had neutral prosodies, most of which referred to the border's geographic position in reference to the United States and/or Mexico. The strongest difference in collocation that can be seen here is the presence of frames determined by the co-occurrence of collocations; for example, while the US media discourse featured BORDER as a sort of 'concept' (which collocated with political LIs and ones related to 'drug-related violence'), FRONTERA was treated like a physical entity within the Mexican corpus. Additionally, the US corpus' use of the node BORDER occurred with both positive and negative collocates (*security*, *shelters*, *drug* and *threatening*) in addition to neutral collocates (i.e. *Mexico*, *years*, *Texas*, etc.), while FRONTERA only collocated with LIs with neutral prosodies.

As in the case of MEXICO, the collocates which occurred with BORDER/FRONTERA also had salient prosodic characteristics which bear mentioning. The collocate *security*, for example, exhibited tendencies toward negative prosody when paired with BORDER, as in *border security*. This was an intriguing example because the pairing of the two LIs cannot be seen as overtly negative (with *border* having neutral prosody and *security* having positive prosody); but, of the five occurrences of 'border security' in the English corpus three occurred in negative contexts. The first use was in reference to the death of an American rancher, the second while discussing so-called 'spillover violence,' and the third occurred as part of a quote implying that a lack of border security is predictive of a lack of

national security in the United States. The concordances of 'border security' are presented in Figure 4, below.

Figure 4. Concordances of *security* with the node word BORDER in the English corpus

at the polls, spent $400 million on BORDER	**security**	and strongly opposed amnesty to people who enter
treating the county as a bellwether of BORDER	**security**	. Indeed, when a Cochise rancher named Robert
immigration by citing his expertise in BORDER	**security**	and spotlighting the potential threat of spillover
have national security until we have BORDER	**security**	," Perry said. The governor also struck that theme
committed "unprecedented resources" to BORDER	**security**	, including a beefed-up Border Patrol and a record

As can be seen in Figure 4, only two occurrences of the collocation of BORDER and *security* did not occur in negative prosodic contexts. Interestingly, both 'non-negative' occurrences still cannot be said to have occurred with overtly positive prosodic characteristics. This was due to the fact that, in the American print media discourse, BORDER *security* would seem to be similar in its prosodic characteristics to *unido* as a collocate of MÉXICO in the Spanish corpus (see Table 12 and discussion, p. 93). That is, security—even when occurring with positive prosodic characteristics—occurs as an opposition to something negative (i.e. the difference in the prosody between something like *security blanket* and *security guard*). Because of these findings, *security* (at least in the context of the corpora examined here) can be seen to have a combined negative prosody due to its co-occurrence with BORDER.

This sort of collocational behavior among seemingly neutral LIs was not only found in the English corpus. In examining the fifth most frequent node in the Spanish corpus PAÍS, a similar pattern was found in which a node which was more frequently used in the overall discourse had a markedly different prosody than did its counterpart in the other corpus.

Table 15. Collocates of PAÍS in the Spanish corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
PAÍS	vive	6	4	7.68
	violento	5	4	7.94
	violencia	59	4	4.38
	méxico	86	4	3.84
	asesinatos	20	4	5.94
	homicidios	19	3	5.60
	tasa	7	2	6.46
	meses	6	2	6.68
	grupos	28	2	4.46

Table 15 shows the significant collocates of the node PAÍS in the Spanish corpus along with the raw frequency, joint frequency and MI score of each. Table 16 presents the same information for the equivalent node COUNTRY in the English corpus.

Table 16. Collocates of COUNTRY in the English corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
COUNTRY	using	7	2	8.68
	uncomfortable	2	2	10.49
	tools	5	2	9.17
	officials	45	2	6.00

While almost half (4 out of 9) of the collocates for the node PAÍS were overtly negative (*violento, violencia, asesinatos, homicidios*), the only overtly negative collocate of the node COUNTRY in the English corpus was *uncomfortable*. Further, many of the apparently positive or neutral collocates of PAÍS carry negative connotations when viewed in the context of their concordances with the node. The collocate *vive,* for example, exclusively occurred in overtly negative contexts within the corpus, with all of its appearances co-occurring with the LI *violencia*.

Figure 5. Concordances of *vive* with the node PAÍS in the Spanish corpus

na dura batalla, pero aún así la violencia que se	**vive**	en muchas partes del PAÍS es enorme. Han salido a
legalizar drogas para erradicar la violencia que	**vive**	el PAÍS El PAN debe entrar a discutir temas noved
men organizado y la situación de violencia que	**vive**	el PAÍS, admitió Fernando Canales Clarion, ex gob
ijo que es necesario que ante la violencia que se	**vive**	en el PAÍS, es ineludible que el tema se discuta

Figure 5 shows the concordance of the LI *vive* within the Spanish corpus. As can be seen, when *vive* was present as a collocate of PAÍS, it exclusively occurred with the LI *violencia*, always as a part of a phrase which roughly translates as 'the violence which X experiences/is experiencing.' This overarching negative semantic prosody was not only found in the co-occurrence patterns of *vive*. The other positive and neutral collocates of PAÍS (*tasa*, *meses* and *grupos)* also acquired negative prosodic features from *homicidios*, *asesinatos* and *drogas* (there was also one occurrence in a context talking about 'cartels'), respectively (see Appendix B, p. 138).

In looking at the examples discussed here (all of which were drawn from the top ten most frequently-used nodes in at least one of the corpora), it becomes clear that there was a general 'pattern' present in both corpora and, by proxy, both countries' print media discourses. That is, in both corpora two general characteristics were seen through this part of the analysis. First, each corpus seemed to have a specific set of discourse characteristics based on the frequently occurring nodes present. Second, these frequent nodes were found to consistently take on negative prosodic features (oftentimes irrespective of the positive or neutral prosody of the individual node or collocate being examined)—see Chapter 5 (p. 115) for further discussion.

Although there were—obviously—instances in which this pattern of negative prosody was not found, many prosodically positive or neutral nodes occurred in overwhelmingly negative prosodic environments; at the same time, most prosodically negative LIs occurred in negative syntactic environments. What was additionally interesting to note in examining the top ten most frequent nodes in each corpus, was the perspective given by these LIs in relation to the overall discourse of each country's media environment and also the 'lived reality' of each country (Berger, 2009). This can be seen very clearly in the focus on the topic of drugs in the US corpus (as seen through the presence of the nodes DRUGS and CARTEL) as well as the presence of BORDER. In the case of the Mexican corpus there was a more marked emphasis on violence and crime. This was not only seen based on the presence of the nodes VIOLENCIA, DROGAS and SECUESTRO, but was also seen in neutral nodes such as GRUPOS and PAÍS which predominately co-occurred with negative prosodic features. As a response to this analysis, which focused mainly on the SP of both corpora insofar as it related to overarching concepts within them, the following section focused more concretely on the shared concepts between both corpora.

In order to accurately analyze the similarities and differences between both corpora's media discourses regarding 'drug-related violence,' it was necessary to focus on the concepts which existed in both corpora. All equivalent concepts were identified in both corpora and were analyzed for their individual prosodic features in much the same way as was done in the preceding section. This, in turn, focused less on the big picture and more on the presentations of concepts which were utilized in both corpora in order to discuss the same topic.

4.2. 'Shared Concept' Nodes

The following section is focused on examining and comparing the use of certain nodes which occurred in both corpora. Whereas the previous section aimed to provide an overview of the discourse characteristics present in the corpora as a whole, the present analysis was focused on nodes which occurred as 'shared concepts'[9] in both corpora. Obviously many of the LIs found in one corpus appeared in the other corpus simply due to the fact that both corpora were built around the shared theme of 'drug-related violence' in Mexico and the United States. Nonetheless, this common discourse was by no means the case in all instances of language use, even in instances where the language used in one corpus seemed logically related to the topic around which the common discourse was built.

All of the following analyses were carried out utilizing the most frequent LIs in both corpora. Within the corpora (115 tokens in the English list and 71 in the Spanish) only thirty tokens occurred as shared concepts across both corpora. This group of shared concepts was analyzed for collocation patterns and general semantic prosody characteristics. This was done (as in the previous section) through the use of the AntConc corpus analysis program (Anthony, 2011). Because of the importance placed on studying concepts and not necessarily individual tokens, the node words were searched here using wildcards (Salama, 2011). This was only done for shared concepts which had different morphological realizations in each corpus. Thus in cases where it was possible, instead of searching DRUG and DRUGS, DRUG* was used as the search term in AntConc and the collocation results brought back were applicable to all appearances of both tokens. This

[9] The term 'shared concept' is used here to refer to nodes in one corpus which had equivalent nodes in the other corpus. 'Shared concepts' is used to describe these nodes because in some cases a literal equivalent was not present and it was thus impossible to examine them as if they were translations of a single meaning.

was not the case in searching for many shared concepts (BORDER and FRONTERA, for example). Due to the fact that the previous section focused on a similar approach, the nodes included here have not already been discussed.

In order to carry out the analysis presented here, all nodes were first sorted in order to assure that only those nodes which were present as concepts in both corpora were examined. This was accomplished by seeking out literal translations whenever possible, but occasionally multiple LIs were included as the shared concept to a LI in the other corpus (this was only the case in the Mexican corpus). For instance, while the node word DRUG had a literal Spanish equivalent in the Mexican corpus (DROGA), a node such as KILLINGS did not have a literal equivalent in the Mexican corpus. In the case of KILLINGS two conceptually equivalent Spanish nodes were found and used as conceptual equivalents (ASESINATOS and HOMICIDIOS). This 'flexibility' in data selection was used because the analysis here was carried out on broad shared concepts and it was, as a result, important to focus on the underlying meanings of the nodes and not on surface similarities (e.g. Saussure's distinction between the signifier and the signified). Being flexible in this sense was very important in examples like that of ASESINATOS and HOMICIDIOS where they are rapidly seen to be two tokens of the same type.

In this way, this section was quite methodologically similar to the majority of SP studies which have been carried out to date. That is, a concept was chosen (here, concepts were 'chosen' based on their presence in both corpora) and the concept's prosody was examined according to its occurrences within the corpora. What sets this analysis apart from many previous SP studies is that prosodies were compared across multi-lingual corpora; while many SP studies have focused on describing the behavior of certain LIs'

prosodies within corpora, the present study sought to describe certain LIs' behavior in the corpus in comparison to another, parallel corpus.

As in the previous section, the first step taken was to look at the overall frequency of positive, negative and neutral nodes present as shared concepts in both corpora. This served to give an insight not only into the common discourse between both corpora but also the 'tone.' As in the previous section, the vast majority of these nodes were neutral. However, in looking only at shared concepts there appeared to be a more varied presence of nodes with positive and negative prosodies than in the previous analysis (Section 4.1, p. 88). This was especially true of nodes with overtly positive prosodies.

The list of shared concepts (see Appendix C, p. 139) contained far more positive nodes than did the list of the ten most frequent occurrences. Additionally, after analyzing the shared concepts it was found that nodes which were seen to be positive or neutral at first glance were not necessarily so depending on the syntactic environment in which they were found. These nodes (POLICE/POLICÍA, AUTHORITIES/AUTORIDADES, MILITARY/EJÉRCITO and NEW/NUEVO) were all deemed to be either positive or neutral. There was really nothing overtly positive or negative about them. That is, none of them immediately suggested a positive prosody like HAPPY would, for example; conversely, none appeared as obviously negative as a LI like WAR would have. Nonetheless, these shared concepts were judged to be more open to carrying a non-neutral connotation (likely a positive one) than would a LI like NORTH, for instance. With this in mind, the following section will first discuss the analysis of these shared concepts before moving on to other nodes which appeared in both corpora.

What was immediately interesting in examining these nodes and their collocates was the steep drop in the number of collocates as compared to the nodes looked at previously. The shared concepts, as a whole, occurred far less frequently than did the nodes examined in Section 4.1 (see p. 88) and, as a result, had a much smaller range of collocates to examine. Despite this, there were many interesting differences to be noted both in terms of collocation and prosodic characteristics. Indeed, in some ways the limited variety of collocates allowed for a more concentrated look at the presence of prosodic features.

In the example of the shared node POLICE/POLICÍA, for example, there was no notable tendency towards either positive or negative prosodic characteristics among the collocates. This was obviously different than what was found in the previous section with the node MEXICO, in which there was a marked presence of negative collocations. However, when comparing the node POLICE to the node POLICÍA, it did become obvious that there were more negative prosodic characteristics in the English corpus than in the Spanish corpus and that even neutral collocates occurred in negative contexts when examined through concordance analysis.

Table 17. Collocates of POLICE in the English corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
POLICE	mexican	61	9	5.59
	federal	33	8	6.31
	officers	22	7	6.70
	state	52	5	4.97
	city	37	5	5.47
	two	34	4	5.27
	municipal	4	4	8.35
	mexico	99	4	3.72
	detained	7	4	7.55
	chief	6	4	7.77
	thursday	13	3	6.24
	luis	10	3	6.62
	local	18	3	5.77
	corruption	4	3	7.94
	allegedly	8	3	6.94

Table 17 shows the most frequently occurring collocates of the node POLICE in the English corpus along with the MI scores and rates of raw and joint frequency for each. Table 18 (below) shows the same information for the node POLICÍA from the Spanish corpus.

Table 18. Collocates of POLICÍA in the Spanish corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
POLICÍA	ministerial	3	3	9.96
	tránsito	3	2	9.38
	san	4	2	8.96
	presentar	2	2	9.96
	nicolás	4	2	8.96
	municipal	4	2	8.96
	federal	9	2	7.79
	división	2	2	9.96
	camioneta	3	2	9.38
	calle	7	2	8.16
	ayer	6	2	8.38
	antidrogas	5	2	8.64

As can be seen in Tables 17 and 18, the node POLICE occurs with one overtly negative collocate in the English corpus (*corruption*), whereas the Spanish node POLICÍA has none. Aside from this, both collocate lists are fairly neutral with both nodes generally collocating with LIs used to differentiate types of police as well as describe law enforcement-related activities. Despite these similarities, when looking at the individual collocates in each list using a concordance analysis tool many differences come to light.

In the case of the node POLICE, the collocates with the highest levels of joint occurrence (*officers*, *municipal*, *detained*, *chief* and *allegedly*) predominately were found in negative prosodic environments despite being (at first glance) LIs with positive or neutral prosodies (see Appendix D, p. 140). For instance, in the case of the collocate *officers*—a seemingly common LI to co-occur with the node POLICE—all but one of the examples found in the text were found to have negative prosodies. Of those, only one was dependent on the greater syntactic context in which it occurred. That is to say that, out of seven co-occurrences with POLICE in the English corpus, the LI *officers* occurred once as a collocate with positive or neutral prosody, once as a potentially neutral collocate within an extremely negative linguistic environment (talking about police officers who had failed background checks) and five times in blatantly negative prosodic environments (allowing violence, being caught with cocaine, committing crimes, etc.).

Figure 6. Concordances of *officers* with the node POLICE in the English corpus

for the Zetas, Domene said. Four POLICE	**officers**	in northern Mexico allowed a violent drug gang to
jail Monterrey, Mexico Several POLICE	**officers**	from Juarez, a suburb of the city of Monterrey,
to San Luis to pay for city POLICE	**officers**	to provide traffic control during periods of long
third of 63,436 low-ranking Mexican POLICE	**officers**	tested so far have failed background and security
percent of midlevel POLICE commanders and	**officers**	. On Thursday, the Defense Department said soldier
Luis Potosi detained two local POLICE	**officers**	from the city of Cardenas with 39 doses of cocaine
who allegedly gave orders to the POLICE	**officers**	to commit illegal acts; documents linking them

Figure 6 shows the concordances of *officers* with the node POLICE in the English corpus. Although the context of many of these occurrences can be seen in the above figure, a more expansive set of concordances for *police officers* can be seen in Appendix D (see p. 140).

This pattern of consistent negative prosody was additionally found when examining the collocate *municipal* (in which the only positive or neutral prosody came about when referencing the killing of six *officers*), *detained* (in which, out of four uses in the corpus, three refer to the *police officers* themselves being detained), *chief** (in which the only non-negative environment in which the collocate was found was in a story saying that a police chief would not lose his job despite having failed a background check[10]) and *allegedly* (which only referred to POLICE being accused of wrongdoing once out of three occurrences in the corpus).

[10] It should be noted here that this particular example is interesting in that it still shows negative prosodic characteristics (albeit not overtly). That is, even though the *chief* did not lose their job they still failed the background check.

As can be seen, the majority of the collocates of POLICE had negative prosodic characteristics in relation to the node. It is worth noting that the collocates which did not exhibit this behavior as strongly were descriptors of the node (*municipal* and *chief**). In both of these cases, there was not a negative prosodic relationship to the node; however, they still occurred in overtly negative syntactic environments in that these examples occurred in relation to the killings and investigations of POLICE. Additionally, the LIs which *did* exhibit negative prosody as related to the node word were LIs which would seem to be common collocates of POLICE in English news writing. However, in the corpus examined here, seemingly common collocates such as *detain** and *alleged** referred to the POLICE themselves and not to the POLICE's actions against criminals or suspected criminals.

Remarkably, this same general presence of negative prosody was not present in the Mexican corpus despite the fact that the POLICE being discussed in the English corpus are the same POLICÍA from the Mexican corpus. That is, in both corpora Mexican law enforcement officials were reported on, not American ones. The most obvious characteristic of the collocates of POLICÍA is that there was not a single overtly negative collocate. Though the English corpus only had one (*corruption*), the collocations found in searching the Mexican corpus were mainly the sort of collocations that one would expect to find in relation to POLICÍA: types of POLICÍA (*ministerial, tránsito, municipal, federal*), time and place vocabulary (*San, Nicolás, calle, ayer, camioneta*) and general descriptive vocabulary related to POLICÍA in terms of police actions and categories (*presentar, división, antidrogas*).

In the case of some of the most strongly collocating LIs—based on MI score (see Appendix D and Appendix E, pp. 140 & 141)—not a single negative prosodic environment was found. That is, while neutral collocates of the English node POLICE took on predominately negative prosodic characteristics due to their surrounding linguistic environments, similar collocates of the Spanish node POLICÍA exhibited no such characteristics. For instance, the conceptual equivalent of the English collocate *municipal*, when examined in the Spanish corpus, revealed positive or neutral connotations in relation to POLICÍA—something not present in the English corpus. Through examining the prosodic characteristics of the collocates of both nodes as well as the syntactic environments in which the nodes were present, it can be gathered that the American press would seem to present a negative image of Mexican POLICE (paradoxically, often portraying them as criminals), while the Mexican press presented Mexican POLICÍA either in a positive light or as victims.

Despite the characteristics detailed above, this sort of imbalance in the distribution of certain connotations was also found in the Mexican corpus when compared to the American corpus. For example, the nodes NEW and NUEVO featured extremely different sets of collocates.

Table 19. Collocates of NEW in the English corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
NEW	york	3	3	10.11
	times	3	2	9.53
	report	7	2	8.31

Table 19 shows the most frequently occurring collocates of the node NEW in the English corpus along with the MI scores and rates of raw and joint frequency for each. Table 20 (below) shows the same information for the node NUEVO in the Spanish corpus.

Table 20. Collocates of NUEVO in the Spanish corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
NUEVO	méxico	86	4	5.80
	león	4	4	10.23
	drogas	57	4	6.39
	tráfico	17	2	7.14
	texas	4	2	9.23
	líder	6	2	8.64
	grupo	12	2	7.64
	empresarios	12	2	7.64
	california	8	2	8.23
	arizona	2	2	10.23

As can be seen in the Tables 19 and 20, the majority of the collocates of NEW/NUEVO in both corpora were focused on place names (New York, Nuevo León, etc.). Nonetheless, in the Mexican corpus there were also two collocates with negative prosodies (*drogas* and *tráfico*) as well as one positive and one neutral collocation (*líder* and *grupo*, respectively) both of which occurred in overtly negative linguistic contexts within the corpus (see Appendix F, p. 142). Thus, as in many of the examples examined above, even positive and neutral collocates seem to acquire negative prosodic characteristics from the contexts in which they appear.

Interestingly, these same sorts of characteristics were found to exist in the shared concepts AUTHORITIES/AUTORIDADES. However, here the general semantic prosody characteristics of the collocates were positive.

Table 21. Collocates of AUTHORITIES in the English corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
AUTHORITIES	mexican	61	4	5.18
	two	34	3	5.61
	cartel	60	3	4.79
	records	6	2	7.53
	helped	8	2	7.11
	federal	33	2	5.07
	drug	108	2	3.36

Table 21 shows the most frequently occurring collocates of the node AUTHORITIES in the English corpus along with the MI scores and rates of raw and joint frequency for each. Table 22 (below) shows the same information for the node AUTORIDADES from the Spanish corpus.

Table 22. Collocates of AUTORIDADES in the Spanish corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
AUTORIDADES	no	64	3	5.33
	mexicanas	4	2	8.74
	acuerdo	12	2	7.16

As can be seen in both Tables 21 and 22, there was a notable difference in the number of collocates for each node. This was, however, most likely due to the frequency of occurrence of the node itself (26 in the English corpus and 14 in the Mexican corpus).

What was noteworthy in examining these shared concepts side by side was that the English collocates of AUTHORITIES generally featured positive prosodic characteristics (as opposed to the prosodic behavior of POLICE). This was even true when looking at overtly negative collocates such as *cartel*. In the case of *cartel*, the LI was found with mostly positive prosodic characteristics (i.e. people associated with a cartel helping AUTHORITIES). In looking at the Spanish corpus, any hard evidence of a marked positive

or negative trend was harder to come by. That is, although all but one collocate were neutral (*not* being negative), the collocates' presence in the corpus was quite mixed.

The final shared concepts which were examined were the nodes MILITARY and EJÉRCITO. In the case of the node MILITARY, nearly all of its occurrences in the English corpus were as parts of quotes from one particular American politician. In the case of EJÉRCITO, the collocates mainly occurred in reference to geographical location (i.e. the Mexican and American militaries) and, likewise, most of the prosodic characteristics were neutral—even in cases where this was not the case, there was by no means a general tendency which could be analyzed and discussed here.

Table 23. Collocates of MILITARY in the English corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
MILITARY	government	33	3	6.55
	support	8	2	8.01
	require	2	2	10.01
	islas	2	2	10.01
	group	19	2	6.76

Table 23 shows the most frequently occurring collocates of the node MILITARY in the English corpus along with the MI scores and rates of raw and joint frequency for each. Table 24 (below) shows the same information for the node EJÉRCITO from the Spanish corpus.

Table 24. Collocates of EJÉRCITO in the Spanish corpus

Node Word	Collocate	Freq.	Joint Freq.	MI
EJÉRCITO	zetas	6	2	8.79
	tecnológicamente	2	2	10.38
	privado	2	2	10.38
	norteamericano	3	2	9.79
	méxico	86	2	4.95
	mexicano	11	2	7.92

While this particular set of shared concepts offered little in the way of easily discernible similarities or differences, it did speak to the general discourse in each corpus with regard to the armed forces. The most obvious example of this is the context in which many of the collocates in Tables 23 and 24 appear. As was mentioned previously, the majority of the collocates in the English corpus came from one particular person whereas the collocates in the Spanish corpus examples were more varied in that they occurred as parts of text as well as parts of quotes. The most notable characteristic which was found in examining MILITARY and EJÉRCITO was based on the personal connotations which appeared to be behind the nodes' presence in both countries' discourses. However, a discussion of this is beyond the scope of the current study and, in many ways, flies in the face of what was attempted in proposing this methodological approach (see Section 5.2, p. 119).

Although the preceding section did not offer concrete examples of marked linguistic differences or similarities between both countries' print media discourses, it did serve as an exploration of the method which was proposed here. In this sense, it would seem that the study was successful in meeting its established goals. The study showed that it is indeed possible to employ a combination of SP and CL methodologies in order to carry out CDA-related research. Additionally, in two cases it was demonstrated that not only was it feasible to carry out research in this way, but that significant findings could come of it.

Part of the reason that the results of the current study are not as well-defined as in other, similar studies in both CDA and SP, is that the current project focused on eliminating the traditional role of the researcher in CDA investigations. That is to say that because the researcher did not explicitly select phenomena, lexical uses or political stances within the corpora from the beginning, it proved difficult to draw marked conclusions about specific phenomena. Despite this, the study presented here was able to show that the methodology

utilized can potentially be used in future research (see Section 5.3, p. 121); and, in fact, individual lexical behaviors were found within the corpora (see the discussions of MEXICO/MÉXICO, POLICE/POLICÍA and BORDER/FRONTERA). The primary reason that this was not more marked, though, was due to various factors involving the methodology and corpora themselves. This is discussed in more detail in the following chapter.

5. Conclusions

Since the main objective of the current project was to separate the basest methodological goals of CDA from the socially- and politically-charged backdrop that has become endemic in carrying out CDA-related research, it is important to discuss the results which came from the methodological steps described above. The general purpose of the current study was to determine whether the methodological approach proposed and utilized here could be a viable methodology for carrying out future CDA research while simultaneously providing a response to some of the criticisms that have been brought up against traditional CDA approaches in the past. Because of this, the discussion presented here consists of three parts, each of which focuses on a different facet of the findings.

With this in mind, the research questions which formed the basis of the project presented here will be addressed in kind. These will then be discussed in more detail in the sections that follow. In response to the first question, the combination of aspects of Methodological Synergy (Baker et al., 2008) and Semantic Prosody was determined to be a feasible option in analyzing texts in two separate languages. In addressing the second question it is important to note that it is impossible to thoroughly and accurately judge the relative objectivity of one's analysis of any findings, especially without something to compare the analysis to. Thus, it seems too early to say whether or not the results of the study were less subjective than traditional CDA research. However, based on the research process and findings here, it can be said that the methodology itself provides a comprehensive, objective approach to the goals of traditional CDA research. Without a doubt there are still improvements to be made, but the methodology was able to respond to

some doubts related to data selection and analysis which have been raised in the past (see, e.g., Mautner, 2009).

The final research question was centered on the potential for future application of the methodology laid out here. It is obviously too early to accurately judge whether the methodology presented above is widely applicable or not. In order to definitively judge practicality in related research, it should first be applied by an outside researcher in order to determine whether or not it is replicable. Nonetheless, when looking at how CDA has progressed from its roots to the present day, it would seem that the future application of this methodology is possible.

The results themselves were broken into two main parts, each focusing on a different way of approaching corpus analysis using the methodology presented above. The first analysis was intended to provide an overview of the language used in each country's print media discourse regarding 'drug-related violence' as well as to highlight any differences or similarities therein. The second part of the corpus analysis was based on a micro-analysis of shared concepts between the corpora. This segment was undertaken with the intent of examining only the discourse common to both corpora in order to examine the differences or similarities present in writing about concepts which were present in both countries. These sections are discussed individually below.

5.1. Frequent Node Analysis

The first set of analyses was based on nothing more than examining the prosodic features of the most common node words in each corpus. As was discussed above, this served the dual purpose of presenting the general language use in each corpus (based on

raw lexical frequency) as well as the general prosodic characteristics of each corpus (based on the prosodic analysis of frequent nodes and their collocates). The results of this analysis were two-fold. That is, both general and specific instances of noteworthy prosodic characteristics were found in examining frequent nodes in both corpora. It bears repeating that the results presented here were representative of the individual features studied. The features studied were all statistically salient based on their occurrence in the corpora (either through examining raw frequency or collocate strength); however, they are only some of the hundreds of features that hypothetically could have been studied.

One of the striking findings related to general prosodic features in both corpora was found in examining the use of the nodes MEXICO and MÉXICO. Although a place name can carry no inherent prosodic value, both corpora seemed to include negative prosodies when Mexico was present in the text. In the case of the English corpus, not a single overtly positive collocate was found to co-occur with MEXICO, and in the Spanish corpus very few positive collocates were found to be present with the node MÉXICO. It should be made clear that many neutral collocates in the English corpus (there were no overtly positive collocates) occurred in negative syntactic environments while in the Spanish corpus both overtly positive and neutral collocates occurred in negative environments. Together these prosodic characteristics would seem to suggest a general negativity portrayed in both countries' print media discourses with regard to Mexico.

Insofar as specific prosodies are concerned, the analysis of the nodes BORDER and FRONTERA highlighted interesting prosodic behaviors, as well as what appeared to be a conceptual difference between both countries' print news media. That is, the US media portrayed the border between both countries as a sort of political boundary as well as a

division between the US and the 'drug-related violence' present in Mexico. This was evident not only based on the high number of prosodically negative collocates (as compared to those found in the Spanish corpus), but also on the collocates themselves. While the English corpus featured many LIs as collocates which focused on the border as a division (*security*, *long*, *state*, *agent*), the Spanish corpus tended toward the use of collocates based on spatial relations (*estados unidos*, *norte*, *ambos*, *lados*). Beyond this use framing in which there are conceptual difference in portrayal, the corpora additionally featured prosodic differences in relation the nodes BORDER and FRONTERA.

This difference was principally centered on the high preference for negative prosodies in the English corpus as compared to the Spanish corpus. This was considered to be of particular relevance due to the fact that not only was the set of English collocates examined found to have many overtly negative LIs, but even those collocates which were deemed to be neutral or positive were found to have overwhelmingly negative prosodic features when viewed in context within the corpus. The most obvious example of this was *security* as a collocate of BORDER (discussed in Section 4.1, p. 88), in which a LI which was seen to have overtly positive prosodic features almost exclusively occurred in negative syntactic environments.

Despite these findings, it is important to consider the fact that the texts analyzed here were representative of a specific discourse as well as a very specific genre of language use. Because of this, it is difficult to assert that these characteristics are representative of a greater discourse. It is entirely possible that the negative prosodic features found in examining otherwise prosodically positive LIs is due to the overall discourse and not

necessarily author intention; however, this differs according to authors' stances (see, e.g., Carvalho, 2008)

5.2. 'Shared Concept' Analysis

The second portion of the corpus analysis was dedicated to the examination of nodes which were present across both corpora. This was done in order to study the differences and similarities in the common discourse between both countries' print news media. This section was based on the perceived necessity of focused analysis from an objective perspective. This was accomplished by avoiding the explicit selection of features to be examined. Nodes in each corpus were paired with their conceptual equivalents (shared concepts) in the opposite corpus and these were then analyzed based on the overt prosodic characteristics which they each possessed. As a result, the shared concepts which were analyzed were those which possessed neither overtly positive nor negative prosodies and which had a high frequency of occurrence in the corpora.

In many ways, this analysis was not groundbreaking in its findings. Nonetheless, some interesting features were noted in comparing the presentations of certain items in both corpora. The clearest example of prosodic contrast found in the shared concepts analysis was that of the nodes POLICE and POLICÍA. This shared concept was found to have a marked difference in prosodic characteristics based on the corpus in which it appeared. While in the Spanish corpus POLICÍA were presented in a generally positive light and as the victims of events, the English corpus presented them in a generally negative light and as criminals. This distinction was seen to be particularly important in that the POLICE/POLICÍA being described in both corpora were Mexican. Because of this there is

obviously the possibility that this prosodic portrayal would change if the nodes referred to American law enforcement personnel.

This striking difference in presentation was further underscored by the analysis of the nodes AUTHORITIES and AUTORIDADES, where there was found to be a marked tendency toward either positive or negative prosody in the Spanish corpus, but in which a positive preference was found in the English data. Referring back to the example of POLICE, though, it is worth noting that in the case of AUTHORITIES and AUTORIDADES, the 'authorities' being referred to were both Mexican and American; thus bolstering the possibility that the nationality of both the 'reporter' and 'reported' may have an influence on prosodic characteristics in a text. However, a definitive response to this possibility is beyond the scope of the current project and would require further research.

Aside from these isolated instances of notable prosodic contrast, the shared concept analysis did not turn up hard evidence of any broad tendencies encompassing *all* of either corpus. Nonetheless, this should not be seen as a disadvantage. The study presented here did in fact find marked prosodic differences that can be seen to be related to the overall discourse in each county's print news media. However, what must be kept in mind is that there was never an expectation to find anything in the analysis (something missing in other, similar studies, which set out to document characteristics already believed to exist). Generally speaking, the present study was based on a corpus approach which was applied to CDA research and not the other way around. As such, it can be inferred that the current study was successful in encountering concrete critical discourse characteristics (most notably in the cases of POLICE/POLICÍA, BORDER/FRONTERA and

AUTHORITIES/AUTORIDADES). However, when compared to similar studies, these findings may seem unimpressive since traditionally CDA research has set out to find what was already assumed to be present (e.g., Salama, 2011) or has found what it 'wanted to' based on the ideologies of researchers (Poole, 2010). Indeed, this is the strength of the current methodology. With no expectations to influence what was examined, the researcher can be almost entirely removed from the analysis process (at least in an ideological sense). Because of this, the present approach—at the very least—could be used to define a general discourse and select items to be analyzed in a traditional sense; but it could also be used as the analysis process itself. In this sense, then, the methodology demonstrated and discussed here can be seen as a viable step forward in improving CDA-based research. Regardless, various factors must be addressed in moving forward. These factors are discussed in the following section.

5.3. Discussion

As is mentioned above, the current study was seen as successful on many fronts. Nonetheless, in order to be able to not only move forward with this style of research but also to rest assured that it is a replicable, viable and accurate approach, three methodological issues must first be addressed. Principle among the issues to be addressed in moving forward is the question of corpus use; namely, is it appropriate to use a comparison corpus, when, and why? Secondly, it is vital to examine the findings themselves in terms of accuracy and implications. Finally, SP research in general is addressed in reference to how it is currently used in research, was used in the present study, and can be used in future studies aiming to carry out CDA-style analyses through the

combination of Methodological Synergy-style analysis (see Salama, 2011; Baker et al., 2008) and Semantic Prosody.

5.3.1. Corpus Issues

The present study was successful in assembling and analyzing representative corpora for a specific language area, and because of this the corpora should be seen as representative of nothing more. The corpora which were analyzed here were 'snapshots' of a single month of texts, from a single type of writing, referencing a single topic and meeting very specific requirements. Therefore, while—in theory—the corpora analyzed here could be seen as representative of greater print media discourse, they should be seen as they are. This is particularly true in that most similar studies have used mega corpora and are thus able to make more generalizable arguments in favor of the applicability of their findings.

Because of this characteristic of the corpora examined here, it cannot be said that the findings laid out above are representative of all uses in the media discourses of either country. This is a particularly important point to address in the future in realtion to the use of Google since it would seem (although it is not explicitly indicated) that Google restricts searches to 160 pages. However at the same time, a comparison of the corpora here and a mega corpus (Salama, 2011; Orpin, 2005; Stubbs, 1997) proves difficult in that a comparison corpus would have to be representative of the same linguistic arena discussed in the present study. In this sense, the use of the corpora should be seen as both an advantage and a disadvantage in terms of methodological characteristics. On the one hand, the present study may not be comparable to other sources and may not be universally

applicable. On the other hand, though, the corpora used here represent a new possibility in CDA research in which micro uses of language can be examined critically without the need for entirely subjective analysis.

5.3.2. Strengths and Weaknesses

Overall, the present study was a success. This was primarily true in that it was capable of responding to many of the issues highlighted in common CDA approaches. That is, data selection was made objective (at least as much so as was possible), research was carried out on a small, practical set of texts and analysis was principally carried out using SP. However, this is not to say that the methodology had no weaknesses. In fact, the approach discussed here had a variety of methodological weaknesses which must be addressed in order to effectively apply the approach in the future. Many of these weaknesses in the methodology have already been mentioned in other parts of the study and so, in the interest of economy, only the most pressing issue will be dissected here.

One of the most far-reaching weaknesses present in the current research was related to the use of SP. Part of the problem in using SP is that it is not terribly common in the literature and that when it does appear it is used as a way to document specific semantic uses, not to measure general discourse characteristics. Because of this, the present use of SP as a methodological tool was novel. The main issue in using SP in an approach like this is the scope of its use. Since it is nearly impossible to examine every use of every node or collocate it becomes difficult to generate concrete findings. One possible solution to this would be to utilize semantic tagging in the corpus analysis; nonetheless, the question of having to assign prosodic value to LIs would remain an issue. While the findings presented

here are certainly concrete insofar as they are true of the language studied, it should be noted that many findings here did not come from the LIs themselves, but rather from their context within the corpora. Because of this, future research may need to take the enormity of some corpora into account in order to completely address CDA issues (similar to Prentice's (2010) use of semantic tagging). A more complete approach in this respect would also help make broader corpus assessments more accurate.

For example, having examined the corpora used in the current study it would seem that the US press utilized more prosodically negative language than did the Mexican press. This is important in that it indicates a non-objective discourse in both countries' coverage. This would be a necessary piece of information in approaching future research because LIs with positive and neutral prosodies were found to very frequently take on negative prosodic characteristics depending on their context of use; thus, it would be very important in future studies to address as many potential prosodic characteristics and behaviors as possible as soon as possible.

Despite this possible methodological weakness, the present study is still seen as having potential for future application. In looking at the current project and the potential applications it has in CDA research, it is important to remember that CDA is interpretative. This interpretation is not in itself bad; and is quite impossible to escape. However, it is possible to improve upon how interpretation is carried out in order to put more rhetorical weight behind any claims made in the field. The difference in this respect between the study presented here and past studies in CDA is that the approach to interpretation, necessary to carry out any CDA study, was changed.

While traditional CDA involves the selection of features to be analyzed by a researcher (based on their personal criteria), the current study relied on the use of statistics to 'select' items for study, and attempted to streamline a form of interpretation which eliminates the use of researchers' personal criteria as much as possible. As a whole, the current study addressed the CDA issues that it set out to, and—at the very least—showed that it is possible to carry out an almost wholly objective study based on some of the goals traditionally held by CDA. Further, the study had many strong points which can hopefully be applied to future CDA studies. Even if only the weak points appear important to the reader, the field will still benefit from the methodology presented here. After all, it was weak points which motivated this study to begin with.

References

Anthony, L. (2011). AntConc (Version 3.2.2) [Computer Software]. Tokyo: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society, 19*(3), 273–306. doi: 10.1177/0957926508088962.

Berger, D. (2009). Constructing crime, framing disaster: Routines of criminalization and crisis in Hurricane Katrina. *Punishment and Society, 11*(4), 491–510. doi: 10.1177/1462474509341139.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, UK: Cambridge University Press.

Blommaert, J., & Bulcaen, C. (2000). Critical discourse analysis. *Annual Review of Anthropology, 29*, 447-466. Retrieved from: EBSCOhost.

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. In: Chiarcos, Eckart de Castilho & Stede (Eds.), *From form to meaning: Processing texts automatically: Proceedings of the Biennial GSCL Conference 2009* (31–40), Tübingen: Gunter Narr Verlag. Retrieved from: http://www.ling.uni-potsdam.de/~gerlof/docs/npmi-pfd.pdf.

Bourdieu, P. (2006). Language and symbolic power. In A. Jaworski & N. Coupland (Eds.), *The discourse reader* (2nd ed.) (506-520). London: Routledge.

Carvalho, A. (2008). Media(ted) discourse and society: Rethinking the framework of critical discourse analysis. *Journalism Studies, 9*(2), 161-177. doi: 10.1080/14616700701848162.

Collins, C., & Jones, P. E. (2006). Analysis of discourse as "a form of history writing": A critique of critical discourse analysis and an illustration of a cultural–historical alternative. *Atlantic Journal of Communication, 14*(1/2), 51-69. Retrieved from: EBSCOhost.

Davis, M. J., & French, T. N. (2008). Blaming victims and survivors: An analysis of post-Katrina print news coverage. *Southern States Communication Journal, 73*(3), 243-257. doi: 10.1080/10417940802219736.

de Beaugrande, R. (2001).Interpreting the discourse of H. G. Widdowson: A corpus-based critical discourse analysis. *Applied Linguistics, 22*(1), 104-121. Retrieved from: University of Wyoming Libraries.

de Beaugrande, R. (1997). The story of discourse analysis. In T. A. van Dijk (Ed.). *Discourse as structure and process* (35-62). London: Sage Publications.

de Swart, H. (1998). *Introduction to natural language semantics*. Stanford, CA: CSLI Publications.

Fairclough, N. (1993). *Discourse and social change*. Cambridge, UK: Polity

Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. London: Longman.

Fairclough, N. (2003). *Analyzing discourse: Textual analysis for social research*. London: Routledge.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. Language, 64(3), 501-538. Retrieved from: JStor.

Freake, R., Gentil, G., & Sheyholislami, J. (2011). A bilingual corpus-assisted discourse study of the construction of nationhood and belonging in Quebec. *Discourse & Society, 22*(1), 21–47. doi: 10.1177/0957926510382842.

Fürsich, E. (2009). In defense of textual analysis: Restoring a challenged method for journalism and media studies. *Journalism Studies, 10*(2), 238-252. doi: 10.1080/14616700802374050.

Glasgow Media Group. (2010). *About the media group*. Retrieved from: http://www.glasgowmediagroup.org/content/section/5/17/.

Google (2011). More search help. Retrieved from: http://support.google.com/websearch/bin/answer.py?hl=en&answer=136861.

Gutiérrez Vidrio, S. (2010). Discurso periodístico: Una propuesta analítica. *Nueva Época, 14*, 169-198. Retrieved from: EBSCOhost.

Haig, E. (2001). Some observations on the critique of critical discourse analysis. *Discourse, 25*(2), 129-149. Retrieved from: http://www.lang.nagoya-u.ac.jp/proj/genbunronshu/25-2/haig.pdf.

Hüning, M. (2012). TextSTAT (Version 2.9) [Computer Software]. Berlin: Freie Universität. Available from http://neon.niederlandistik.fu-berlin.de/en/textstat/.

Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics, 12*(2), 249–268. Retrieved from: EBSCOhost.

Jones, P. E. (2004). Discourse and the materialist conception of history: Critical comments on critical discourse analysis. *Historical Materialism, 12*(1), 97-125. Retrieved from EBSCOhost.

Jones, P. E., & Collins, C. (2006). Political analysis versus critical discourse analysis in the treatment of ideology: Some implications for the study of communication. *Atlantic Journal of Communication, 14*(1/2), 28-50. Retrieved from: EBSCOhost.

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language, 39*(2), 170-210.

Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.

Leech, G. (1981). *Semantics: The study of meaning* (2nd ed.). Middlesex, England: Penguin Books.

Lewis, S. C., & Reese, S. D. (2009). What is the war on terror? Framing through the eyes of journalists. J&MC Quarterly, 86(1), 85-102. Retrieved from: EBSCOhost.

Louw, B. (2008). Contextual prosodic theory: Bringing semantic prosodies to life. *Texto!, 8*(1), 1-58. Retrieved from: http://www.revue-texto.net/docannexe/file/124/louw_prosodie.pdf.

Mautner, G. (2009). Checks and balances: How corpus linguistics can contribute to CDA. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (2nd ed.) (122-143). Los Angeles: Sage.

Morley, J., & Partington, A. (2009). A few frequently asked questions about semantic—or evaluative—prosody. *International Journal of Corpus Linguistics, 14*(2), 139-158. doi: 10.1075/ijcl.14.2.01mor.

Orpin, D. (2005). Corpus linguistics and critical discourse analysis: Examining the ideology of sleaze. *International Journal of Corpus Linguistics, 10*(1), 37-61. Retrieved from: EBSCOhost.

Orwell, G. (1946). Politics and the English language. Retrieved from: http://www.mtholyoke.edu/acad/intrel/orwell46.htm

Oster, U. (2010). Using corpus methodology for semantic and pragmatic analyses: What can corpora tell us about the linguistic expression of emotions?. *Cognitive Linguistics, 21*(4), 727–763. doi: 10.1515/COGL.2010.023.

Partington, A. (2007). Irony and reversal of evaluation. *Journal of Pragmatics, 39*, 1547–1569. doi:10.1016/j.pragma.2007.04.009.

Partington, A. (2004). “Utterly content in each other’s company”: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics, 9*(1), 131-156. Retrieved from: EBSCOhost.

Peñaranda Contreras, L. (2010). De ‘marijuaneros’ a ‘mafiosos’: Transformaciones de los discursos de la prensa colombiana de los años sesenta y setenta sobre el ‘problema droga’. *Signo y Pensamiento, 57*(XXIX), 314-329. Retrieved from: EBSCOhost.

Ping-Fang, Y., & Jing-Chun, C. (2009). Semantic prosody: A new perspective on lexicography. *US-China Foreign Language, 7*(1), 20-25. Retrieved from: EBSCOhost.

Poole, B. (2010). Commitment and criticality: Fairclough’s critical discourse analysis evaluated. *International Journal of Applied Linguistics, 20*(2), 137-155. Retrieved from: EBSCOhost.

Prentice, S. (2010). Using automated semantic tagging in Critical Discourse Analysis: A case study on Scottish independence from a Scottish nationalist perspective. *Discourse & Society, 21*(4), 405–437. doi: 10.1177/0957926510366198.

Salama, A. H. Y. (2011). Ideological collocation and the recontexualization of Wahhabi-Saudi Islam post-9/11: A synergy of corpus linguistics and critical discourse analysis. *Discourse & Society, 22*(3), 315–342. doi: 10.1177/0957926510395445.

Schrøder, K. C. (2007). Media discourse analysis: Researching cultural meanings from inception to reception. *Textual Cultures: Texts, Contexts, Interpretation, 2*(2), 77-99. doi: 10.1353/txc.0.0077.

Sinclair, J. (1991). *Corpus concordance collocation*. Oxford, UK: Oxford University Press.

Stubbs, M. (2007). Colloction and semantic profiles: On the cause of the trouble with quantitative studies. In W. Teubert & R. Krishnamurthy (Eds.) *Corpus linguistics: Critical concepts in linguistics* (1-24). London: Routledge.

Stubbs, M. (2005). Conrad in the computer: Examples of quantitative stylistic methods. *Language and Literature, 14*(1), 5-24. Retrieved from: EBSCOhost.

Stubbs, M. (2001a).Texts, corpora, and problems of interpretation: A response to Widdowson. *Applied Linguistics, 22*(2), 149-172. Retrieved from: University of Wyoming Libraries.

Stubbs, M. (2001b). On inference theories and code theories: Corpus evidence for semantic schemas. *Text, 21*(3), 437-465. Retrieved from: EBSCOhost.

Stubbs, M. (1997). Whorf's children: Critical comments on Critical Discourse Analysis (CDA). In Ann Ryan & Alison Wray (Eds.), *Evolving models of language* (1-15). Clevedon, UK: Multilingual Matters.

Törrönen, J. (2004). Finnish newspapers' positions on drug policy between 1993 and 2000. Contemporary Drug Problems, 31, 59-88. Retrieved from: EBSCOhost.

Túñez, M., & Guevara, M. (2009). Framing por proximidad como criterio de noticiabilidad: La curva de las ausencias. *Revista Latina de Comunicación Social, 64*, 1030-1044. doi: 10.4185/RLCS-64-2009-878-1.030-1.044.

Van Dijk, T. (2006). Discourse and the denial of racism. In A. Jaworski & N. Coupland (Eds.), *The discourse reader* (2nd ed.) (506-520). London: Routledge.

Van Dijk, T. A. (1997). The study of discourse. In T. A. van Dijk (Ed.). *Discourse as structure and process* (1-34). London: Sage Publications.

Van Dijk, T. A. (1988). Texto y contexto: Semántica y pragmática del discurso (3rd ed.). (J. D. Moyano, Tran.). Madrid: Catedra.

van Hout, T., & MacGilchrist, F. (2010). Framing the news: An ethnographic view of business newswriting. *Text & Talk, 30*(2), 169–191. doi: 10.1515/TEXT.2010.009.

Weiss, G., & Wodak, R. (2003). Introduction: Theory, interdisciplinarity and critical discourse analysis. In G. Weiss & R. Wodak (Eds.). *Critical discourse analysis: Theory and interdisciplinarity* (1-34). New York: Palgrave Macmillan.

Whitsitt, S. (2005). A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics, 10*(3), 283-305. Retrieved from: EBSCOhost.

Widdowson, H. G. (2001a). Scoring points by critical analysis: A reaction to Beaugrande. *Applied Linguistics, 22*(2), 266-272. Retrieved from: University of Wyoming Libraries.

Widdowson, H. G. (2001b). Interpretations and correlations: A reply to Stubbs. *Applied Linguistics, 22*(4), 531-538. Retrieved from: University of Wyoming Libraries.

Wodak, R., & Meyer, M. (2009). Critical discourse analysis: History, agenda, theory and methodology. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (2nd ed.) (2-33). Los Angeles: Sage.

Wodak, R. (2007). Pragmatics and critical discourse analysis: A cross-disciplinary inquiry. *Pragmatics and Cognition, 15*(1), 203-225. Retrieved from: www.ling.lancs.ac.uk/staff/wodak/papers/wodak_pragdiscourse.pdf.

Wodak, R. (2000). Critical discourse analysis at the end of the 20th century. *Research on Language and Social Interaction, 32*(1&2), 185–193. Retrieved from: EBSCOhost.

Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics, 27*(1), 103–129. doi:10.1093/applin/ami045.

Yusuf, Y. Q. (2010). A corpus-based linguistic analysis on spoken corpus: Semantic prosodies on “robots”. Journal of Language and Linguistic Studies, 6(1), 49-63. Retrieved from: EBSCOhost.

Zillmann, D., Chen, L., Knobloch, S., & Callison, C. (2004). Effects of lead framing on selective exposure to internet news reports. *Communication Research, 31*(1), 58-81. doi: 10.1177/0093650203260201.

Zhang, C. (2010). An overview of corpus-based studies of semantic prosody. *Asian Social Science, 6*(6), 190-194. Retrieved from: EBSCOhost.

Zhang, W. (2009). Semantic prosody and ESL/EFL vocabulary pedagogy. *TESL Canada Journal, 26*(2), 1-12. Retrieved from: EBSCOhost.

Zillmann, D., Chen, L., Knobloch, S., & Callison, C. (2004). Effects of lead framing on selective exposure to internet news reports. *Communication Research, 31*(1), 58-81. doi: 10.1177/0093650203260201.

Appendix A

English Corpus	Spanish Corpus
DRUG	MÉXICO
MEXICO	VIOLENCIA
BORDER	DROGAS
MEXICAN	ESTADOS
CARTEL	PAÍS
STATE	UNIDOS
OFFICIALS	SECUESTRO
POLICE	GRUPOS
ONE	AÑOS
VERACRUZ	ESTADO
VIOLENCE	MEXICANOS
CITY	DOS
ENFORCEMENT	DROGA
STATES	SEGURIDAD
ZETAS	CÁRTELES
TWO	CONTRA
FEDERAL	ASESINATOS
GOVERNMENT	GOBIERNO
PERRY	HOMICIDIOS
UNITED	JUÁREZ
LAW	CIUDAD
SECURITY	NARCOTRÁFICO
TEXAS	TRÁFICO
PEOPLE	AÑO
AUTHORITIES	CHIHUAHUA
FOUND	MUERTOS
ILLEGAL	ACAPULCO
AMERICAN	AUTORIDADES
CARTELS	DÍA
INFORMANT	HAY
OFFICERS	MAYOR
AGENTS	PERSONAS
DRUGS	CIUDADES
IMMIGRATION	CÁRTEL
GROUP	HERNÁNDEZ
YEARS	NACIONAL
CRIME	PARTE
IMMIGRANTS	SINALOA
LOCAL	UNO

ARBABSIAR	ACUERDO
BODIES	CRIMEN
INFORMATION	EMPRESARIOS
JUST	FRONTERA
MEMBERS	GRUPO
MEN	POLICÍA
SINALOA	SITUACIÓN
WEAPONS	TRES
WEEK	CARTELES
ANONYMOUS	GARZA
CALDERON	MEXICANO
EL	ORGANIZACIÓN
FAST	ORGANIZADO
FURIOUS	CASOS
KNOW	CONSUMO
LAST	GUERRERO
RECENT	HOMBRE
BACK	NUEVO
PLOT	POLÍTICA
VIOLENT	TEMA
WAR	CIENTO
ADMINISTRATION	COCAÍNA
AMBASSADOR	DELITO
FAR	DELITOS
KILLED	DINERO
MILITARY	EJÉRCITO
MONTH	FAMILIA
PUBLIC	FEDERAL
SAN	GENERAL
SEVERAL	NORTE
INFORMANTS	TIENE
INTELLIGENCE	ZONA
IRANIAN	
NEW	
PLAN	
POLITICAL	
PRINCE	
THURSDAY	
TOP	
AGENT	
ALLEGED	
IRAN	
KILLINGS	

OPERATION PRESIDENT THREE TRAFFICKING WORK CONTROL DANGEROUS DEPARTMENT DPS FIRST FORMER GANG HOUSE NORTHERN PROGRAM YEAR COUNTRY FORCES GANGS GROUPS GUNS HOME INCLUDING INTERNATIONAL KILL LUIS MAKE NATIONAL NORTH PASO REPUBLICAN SAUDI TROOPS	

Appendix B

Concordances for collocates of PAÍS:

tasa

se reportó un aumento de 65 por ciento en la	**tasa**	de homicidios del PAÍS. Asegura que debido a esta
habitantes, mientras que en el resto del PAÍS la	**tasa**	era de 507. En el 2010 el promedio de homicidios

meses

violento del PAÍS. Más de 1,500 asesinatos en 20	**meses**	como consecuencia delos enfrentamientos entre "L
violento del PAÍS. Más de 1,500 asesinatos en 20	**meses**	como consecuencia de los enfrentamientos entre "L

grupos

ntes que existían en ese PAÍS, haciendo que estos	**grupos**	buscaran otros lugares y parece ser que lo encont
en todo el PAÍS. Modalidad realizada por ciertos	**grupos**	del trasiego de drogas. – Secuestro exprés: Plag

Appendix C

List of shared concept nodes[11]

English Corpus	**Spanish Corpus**
DRUG*	DROGA*
MEXIC*	MÉXIC* (MEXIC*)
BORDER	FRONTERA
VIOLEN*	VIOLEN*
STATE*	ESTAD*
CARTEL*	CÁRTEL* (CARTEL*)
COUNTR*	PAÍS (PAISES)
UNITE*	UNIDO*
POLICE	POLICÍA
GROUP*	GRUPO*
YEAR*	AÑO*
ONE	UNO
CIT*	CIUDAD*
TWO	DOS
SECURITY	SEGURIDAD
FEDERAL	FEDERAL*
KILLINGS	ASESINATOS (HOMICIDIOS)
GOVERNMENT	GOBIERNO
TRAFFICKING	TRÁFICO
AUTHORITIES	AUTORIDADES
PEOPLE	PERSONAS
SINALOA	SINALOA
CRIME*	CRIMEN* (DELITO*)
NATIONAL	NACIONAL
THREE	TRES
MEN	HOMBRE
MILITARY	EJÉRCITO
NEW	NUEVO
POLITICAL	POLÍTICA
NORTH*	NORTE

[11] LIs which appear in parenthesis in the Spanish corpus are shared concepts as well. These were examples in which a simple literal equivalent would have excluded a conceptually equivalent LI.

Appendix D

Collocates for the node POLICE

Officers

MONTERREY, Mexico Several POLICE	**officers**	in northern Mexico allowed a violent drug gang to
for the Zetas, Domene said. Four POLICE	**officers**	from Juarez, a suburb of the city of Monterrey,
Stonegarden to San Luis to pay for city POLICE	**officers**	to provide traffic control during periods of long
third of 63,436 low-ranking Mexican POLICE	**officers**	tested so far have failed background and security
percent of midlevel POLICE commanders and	**officers**	. On Thursday, the Defense Department said soldier
Luis Potosi detained two local POLICE	**officers**	from the city of Cardenas with 39 doses of cocaine
allegedly gave orders to the POLICE	**officers**	to commit illegal acts; documents linking them to

Municipal

eral forces would put 21,035 state and	**municipal**	POLICE in Veracruz under scrutiny and purge
sing since Monday. So far in 2011, six	**municipal**	POLICE chiefs have been killed in Michoacan,
terrey said he was going to keep the	**municipal**	POLICE chief and the director of the municipal
POLICE chief and the director of the	**municipal**	POLICE monitoring center in their jobs, even

Detained

federal POLICE said Thursday they had	**detained**	two U.S. men for attempting to fly out of Mexico
in the northern state of San Luis Potosi	**detained**	two local POLICE officers from the city of Carden
measured as a gram. The soldiers then	**detained**	the city's POLICE chief and a police commander,

Chief*

-ridden body of Javier Luis Mejia, the POLICE	**chief**	of the small Michoacan town of Aporo.
Monday. So far in 2011, six municipal POLICE	**chiefs**	have been killed in Michoacan, a state dominate
aid he was going to keep the municipal POLICE	**chief**	and the director of the municipal POLICE
urity checks. Almost one-quarter of the POLICE	**chiefs**	and top commanders tested so far have also
The soldiers then detained the city's POLICE	**chief**	and a POLICE commander, who allegedly gave

Allegedly

POLICE said they detained two U.S. citizens for	**allegedly**	carrying two metric tons of marijuana in
of taxi drivers, POLICE, journalists and others	**allegedly**	working with the Zetas. He did not say how
POLICE chief and a POLICE commander, who	**allegedly**	gave orders to the POLICE officers to

Appendix E

Collocates for the node POLICÍA

Ministerial

la batea de una camioneta de la POLICÍA	**Ministerial**	, estacionada en la calle Tabachines, a unos
la batea de una camioneta de la POLICÍA	**Ministerial**	había un hombre muerto, por lo que dio parte a
batea de la patrulla 01284 de la POLICÍA	**Ministerial**	. Asimismo, en el parabrisas del lado izquierdo

Tránsito

que un POLICÍA y un elemento de	**tránsito**	resultaron lesionados a balazos en dos hechos dis
Valenzuela Hernández, POLICÍA de	**tránsito**	de San Nicolás de los Garza fue atacado a balazos

Municipal

una amenaza a la POLICÍA preventiva	**municipal**	de Emiliano Zapata, por lo que se ha tratado
de manera mensual realiza la POLICÍA	**Municipal**	tanto en infractores como presuntos

Appendix F

Collocates for the node NUEVO

líder

por ser muy delicado. MÉXICO, el nuevo	**líder**	del tráfico de drogas México es el nuevo líder en
tráfico de drogas MÉXICO es el nuevo	**líder**	en el tráfico de drogas, desplazando a Colombia.
en agosto del 2010. Así llegó a ser el	**líder**	de Acapulco, aunque operaba desde el Estado de MÉXIC

grupos

Estados Unidos, alerta sobre el crecimiento de	**grupos**	paramilitares en MÉXICO, algunos de los cuales
incluso retar al Estado. Al respecto, refiere que	**grupos**	paramilitares de MÉXICO son sospechosos de
la situación en MÉXICO y del surgimiento de	**grupos**	montar paramilitares, e incluso ha planteado que
no tripulados al espacio aéreo mexicano. “Los	**grupos**	del crimen organizado en MÉXICO se han

drogas

México, el NUEVO líder del tráfico de	**drogas**	México es el nuevo líder en el tráfico de drogas,
México es el NUEVO líder en el tráfico de	**drogas**	, desplazando a Colombia. Colombia combatió con
abra el debate de la legalización o no de las	**drogas**	. Empresarios de NUEVO León e integrantes de
empresarios de NL por legalización de	**drogas**	Un grupo de empresarios de NUEVO León

tráfico

delicado”. México, el NUEVO líder del	**tráfico**	de drogas México es el nuevo líder en el tráfico
drogas México es el NUEVO líder en el	**tráfico**	de drogas, desplazando a Colombia. Colombia