

**METODOLOGÍA PARA LA CARACTERIZACIÓN DE PATRONES  
[VERBO + OBJETO DIRECTO] SOBRE POBREZA EN LA PRENSA  
COLOMBIANA PARA PROPÓSITOS DE ANÁLISIS DE  
SENTIMIENTOS**

**CARLOS MARIO PÉREZ PÉREZ**

Trabajo de investigación presentado para optar al título de:  
**MAGÍSTER EN LINGÜÍSTICA**

Asesor

**GABRIEL ÁNGEL QUIROZ HERRERA**

Doctor en Lingüística Aplicada

**MAESTRÍA EN LINGÜÍSTICA  
FACULTAD DE COMUNICACIONES  
UNIVERSIDAD DE ANTIOQUIA**

**2018**



**UNIVERSIDAD  
DE ANTIOQUIA**

1 8 0 3

Del Evangelio según San Juan, "En el principio existía la Palabra y la Palabra estaba con Dios, y la Palabra era Dios." (Juan, 1-1). Ahora bien, en Antonio de Nebrija, el verbo es por excelencia la palabra; entonces, siguiendo la tradición Latina...

*"En el principio existía el Verbo y el Verbo estaba con Dios, y el Verbo era Dios."* (Juan, 1-1).

A mis padres, Aleida María y Luis Emilio.

A mi abuela, María del Consuelo.

A mi tía Viviana y a Tato.

A mis hermanos, Julián Andrés, María Isabel y Samuel.

**Los frutos de este trabajo de investigación siempre serán cosechados en familia.**

## AGRADECIMIENTOS

En primer lugar, gracias a Dios por el don de la vida, por las bendiciones recibidas y por permitir que este proyecto de investigación llegue a buen puerto.

Gracias a mi familia, quienes vivieron de cerca las altas y bajas durante el transcurso de estos dos años de trabajo. Su motivación ha sido el motor fundamental para alcanzar la meta. Los frutos de esta investigación siempre serán cosechados en familia.

Profesor Gabriel, gracias por la paciencia, la dedicación y las enseñanzas; por las largas jornadas de trabajo y por la confianza en el desarrollo de esta investigación. Gracias por todo el conocimiento ofrecido, nunca hubo una pregunta sin respuesta. Gracias por las palabras de ánimo y tranquilidad en los momentos en que se nubla el horizonte. He tenido la dicha de contar con un asesor de primer nivel.

A todos los miembros del grupo de investigación de Traducción y Nuevas Tecnologías (TNT) de la Universidad de Antioquia. Gracias a ustedes he confirmado el valor de compartir el conocimiento. Las múltiples conversaciones, llenas de contenido, permitieron refinar esta investigación. Gracias por sus valiosos aportes en los momentos precisos. En especial, a los profesores Giovanni Diaz, Antonio Tamayo, José Luis Rojas y Carlos Muñoz, por dedicar su tiempo en la validación de los datos o en la lectura del escrito final. Sus precisiones han sido vitales en la culminación de esta investigación.

A mis amigos, gracias por el apoyo constante, las palabras de aliento, los buenos deseos. Gracias por la compañía a lo largo de este recorrido. Jessica y Mario, todo lo que hay por decirles, ya se los he expresado. Gracias a ambos.

Wilfer y Juan, las extensas jornadas de trabajo entre aciertos y dudas, cansancios y risas; han permitido que este trabajo vea su final. Gracias por la presión, apoyo y acompañamiento en la etapa final de escritura de esta investigación. Buen viento y buena mar en los proyectos futuros.

Finalmente, gracias a la Universidad de Antioquia y a la Facultad de Comunicaciones, por los espacios de formación brindados. A los profesores de la maestría en Lingüística, gracias por permitirme crecer como lingüista.

## RESUMEN

El Análisis de Sentimientos permite la clasificación de la polaridad de textos en positivos, negativos o neutros al combinar las ventajas de la lingüística aplicada y las ciencias computacionales. Desde el Procesamiento del Lenguaje Natural (PLN), la construcción de herramientas para el análisis de sentimientos ha ocupado un espacio importante en las investigaciones en el área, desde el año 2000, con diversos trabajos en distintos niveles de análisis (oración, entidad-aspecto, documento), donde se ha privilegiado el adjetivo como principal portador de la polaridad relegando a un segundo plano las demás categorías gramaticales. Adicionalmente, para el español, existen pocas herramientas, lo que ha dificultado un mayor avance en el desarrollo de aplicaciones para clasificar la polaridad de los textos en esta lengua. Esta investigación tiene como objetivo describir las ocurrencias del patrón [verbo + objeto directo], extraídas de las secciones de política y economía sobre pobreza en tres periódicos colombianos (El Colombiano, El Tiempo y El Espectador); como elementos portadores de polaridad, discriminantes en la construcción de herramientas en español para el análisis de sentimientos y capaces de establecer los referenciales mediáticos sobre el tema en Colombia; a través de un enfoque basado en la lingüística de corpus. Los resultados sugieren que una correcta clasificación de las ocurrencias del patrón [verbo + objeto directo] (las cuales aportarían polaridad como los adjetivos), podría permitir el mejoramiento los resultados en la clasificación de la polaridad de los textos en español, al ser utilizados en aplicaciones de análisis de sentimientos. Así mismo, a partir de la interpretación de los datos, se podría establecer cómo se presenta el fenómeno de la pobreza en Colombia en los periódicos trabajados, los cuales establecen la agenda económica y política del país. Por último, esta investigación constituye un avance en la solución a los desafíos inherentes al desarrollo de herramientas para Análisis de Sentimientos en español, en cuanto se presenta una guía metodológica para clasificar este tipo de patrón [verbo + objeto directo].

Palabras Clave: *Análisis de sentimientos, verbo, complemento de objeto directo.*

## ABSTRACT

Sentiment analysis deals with the classification of text according to their polarity as positive, negative or neutral in different levels (sentence, entity-aspect, document) by implementing Natural Language Processing (PNL) techniques along with Applied Linguistics bases. Since 2000, due to the spread of media information generated through social networks, researches have been developing applications on detecting the strength of subjectivity of a text. To date, there is almost no work on the use of verbs on this field nor has been any work on Spanish applications. This research aims at classifying the entries [verb + direct object], extracted from political and social news on poverty in Colombian newspapers. We propose a classification of these entries based on syntactic and semantic criteria, in order to establish the discriminant features and to design a methodological guide for sentiment analysis. We describe the results of experiments on a set of 1,360 entries [verb + direct object], retrieved from 205 news articles, corresponding to 50 % of the total entries, after the data cleaning process. Our results suggest that an accurate classification of these entries might allow a better performance of the sentiment analysis algorithms and also might help on the establishment of the media references on poverty in Colombian newspapers.

Keywords: *Sentiment Analysis, Verb, Direct Object Complement.*

# TABLA DE CONTENIDO

AGRADECIMIENTOS .....	IV
RESUMEN.....	V
ABSTRACT.....	VI
LISTA DE FIGURAS.....	XIV
LISTA DE TABLAS .....	XVI
LISTA DE ANEXOS .....	XVIII
<b>PARTE I: INTRODUCCIÓN, OBJETIVOS Y ESBOZO DEL TRABAJO DE INVESTIGACIÓN.....</b>	<b>1</b>
CAPÍTULO 1: INTRODUCCIÓN.....	2
1.1. <i>HIPÓTESIS</i> .....	3
1.2. <i>OBJETIVOS</i> .....	4
1.2.1.    GENERAL.....	4
1.2.2.    ESPECÍFICOS.....	4
1.3. <i>ESBOZO DEL TRABAJO DE INVESTIGACIÓN</i> .....	5
<b>PARTE II: ANÁLISIS DE SENTIMIENTOS Y EL VERBO EN ESPAÑOL.....</b>	<b>6</b>
CAPÍTULO 2: ANÁLISIS DE SENTIMIENTOS.....	7
2.1. <i>¿QUÉ ES EL ANÁLISIS DE SENTIMIENTOS?</i> .....	7
2.2. <i>NIVELES DE ANÁLISIS</i> .....	8
2.2.1.    ANÁLISIS DE SENTIMIENTOS A NIVEL DE DOCUMENTO.....	8
2.2.2.    ANÁLISIS DE SENTIMIENTOS A NIVEL DE ORACIÓN.....	8
2.2.3.    ANÁLISIS DE SENTIMIENTOS A NIVEL DE ENTIDAD-ASPECTO.....	9
2.3. <i>APLICACIONES DEL ANÁLISIS DE SENTIMIENTOS</i> .....	10
2.4. <i>EL ANÁLISIS DE SENTIMIENTOS DESDE DIFERENTES PERSPECTIVAS</i> .....	11
2.4.1.    EL ANÁLISIS DE SENTIMIENTOS DESDE LA PERSPECTIVA LINGÜÍSTICA.....	11
2.4.1.1.    PRESENCIA DE TÉRMINO VS. FRECUENCIA.....	11
2.4.1.2.    CARACTERÍSTICAS BASADAS EN TÉRMINOS MÁS ALLÁ DE LOS UNIGRAMAS.....	12
2.4.1.3.    INFORMACIÓN DE LAS CATEGORÍAS GRAMATICALES ( <i>PART-OF-SPEECH INFORMATION</i> , POS).....	13

2.4.1.4.	SINTAXIS .....	13
□	LA NEGACIÓN.....	14
2.4.1.5.	CARACTERÍSTICAS ORIENTADAS A TEMAS ( <i>TOPIC-ORIENTED FEATURES</i> ) .....	15
2.4.2.	ANÁLISIS DE SENTIMIENTOS DESDE LA PERSPECTIVA COMPUTACIONAL.....	15
2.4.2.1.	ANÁLISIS DE SENTIMIENTOS CON MODELOS DE TÓPICOS .....	15
□	ANÁLISIS DE SENTIMIENTOS A PARTIR DEL APRENDIZAJE NO SUPERVISADO Y APRENDIZAJE SUPERVISADO.....	16
CAPÍTULO 3: DEL VERBO EN ESPAÑOL PARA PROPÓSITOS DE ANÁLISIS DE SENTIMIENTOS .....		17
3.1.	EL VERBO EN ESPAÑOL PARA PROPÓSITOS DE ANÁLISIS DE SENTIMIENTOS .....	18
3.1.1.	EL VERBO EN LOS NIVELES DE LA LENGUA .....	22
3.1.1.1.	MORFOLOGÍA .....	22
3.1.1.1.	PERSONA Y NÚMERO .....	22
3.1.1.1.1.	TIEMPO .....	23
3.1.1.1.2.	MODO.....	23
3.1.1.1.3.	ASPECTO .....	24
3.1.1.2.	SINTAXIS .....	25
3.1.1.2.1.	TEORÍA DE LAS VALENCIAS VERBALES.....	27
3.1.1.3.	SEMÁNTICA .....	28
3.1.1.3.1.	TIPOLOGÍA VERBAL.....	29
3.1.1.3.1.1.	LEVIN (1993) .....	29
3.1.1.3.1.2.	CLASIFICACIÓN VERBAL EN WORDNET .....	30
3.1.1.3.1.3.	LORENTE (2002) .....	33
3.1.1.3.1.3.1.	VERBOS DISCURSIVOS .....	35
3.1.1.3.1.3.2.	VERBOS CONECTORES .....	35
3.1.1.3.1.3.3.	VERBOS FRASEOLÓGICOS .....	36
3.1.1.3.1.3.4.	VERBOS-TÉRMINO .....	36



3.1.1.3.1.4. OTRAS PROPUESTAS DE CLASIFICACIÓN.....	36
3.1.2. COMPLEMENTOS ARGUMENTALES.....	37
3.1.2.1. TIPOS DE COMPLEMENTOS ARGUMENTALES.....	38
3.1.2.1.1. COMPLEMENTOS NUCLEARES Y MARGINALES .....	38
3.1.2.1.2. COMPLEMENTOS ARGUMENTALES O REGIDOS Y NO ARGUMENTALES .....	39
3.1.2.1.3. COMPLEMENTOS INTEGRABLES Y NO INTEGRABLES .....	39
3.1.2.2. COMPLEMENTO DIRECTO O IMPLEMENTO .....	40
3.1.2.3. OTROS TIPOS DE COMPLEMENTOS: INDIRECTO, SUPLEMENTO Y AGENTE.....	44
3.1.2.3.1. COMPLEMENTO INDIRECTO .....	44
3.1.2.3.2. SUPLEMENTO Y COMPLEMENTO AGENTE.....	44
<b>PARTE II: METODOLOGÍA DE INVESTIGACIÓN .....</b>	<b>46</b>
CAPÍTULO 4: METODOLOGÍA DE INVESTIGACIÓN .....	47
4.1. ENFOQUE METODOLÓGICO.....	47
4.2. ALCANCE DE LA INVESTIGACIÓN.....	48
4.3. ENFOQUE BASADO EN CORPUS (CORPUS- BASED).....	48
4.4. CONCEPCIÓN DEL CORPUS DE TRABAJO Y METODOLOGÍA DE TRABAJO .....	49
4.4.1. DESCRIPCIÓN DEL CORPUS DE TRABAJO .....	49
4.4.2. DESCRIPCIÓN DE LA MUESTRA DE TRABAJO .....	51
4.5. HERRAMIENTAS Y RECURSOS.....	53
4.5.1. HERRAMIENTAS PARA EL ETIQUETAJE DE LOS DATOS.....	54
4.5.1.1. WORDNET 2.1.....	54
4.5.1.2. DICCIONARIO COMBINATORIO PRÁCTICO DEL ESPAÑOL CONTEMPORANEO.....	55
4.5.1.3. DICCIONARIO DE LA REAL ACADEMIA DE LA LENGUA ESPAÑOLA.....	55
4.6. METODOLOGÍA .....	56
4.7. ETIQUETAJE.....	57
4.7.1. NÚCLEO DEL OBJETO DIRECTO .....	59

4.7.2.	ÁREA TEMÁTICA DEL OBJETO DIRECTO .....	59
4.7.2.1.	GENERAL .....	60
4.7.2.2.	ECONOMÍA .....	60
4.7.2.3.	POLÍTICA .....	60
4.7.2.4.	SOCIEDAD.....	60
4.7.3.	TIPO DE VERBO .....	61
4.7.4.	SENTIDO DEL VERBO.....	61
4.7.5.	CARGAS DE LOS VERBOS Y DE LOS OBJETOS DIRECTOS .....	62
4.8.	ESTADÍSTICA.....	62
CAPÍTULO 5: METODOLOGÍA PARA LA CONSTRUCCIÓN DE UNA BASE DE DATOS [VERBO + OBJETO DIRECTO] .....		63
5.1.	ETAPA 1: EXTRACCIÓN DE PATRONES [VERBO + OBJETO DIRECTO] .....	63
5.2.	ETAPA 2: DEPURACIÓN DE LAS OCURRENCIAS EXTRAÍDAS.....	64
5.2.1	ELIMINACIÓN DE ERRORES EN LA EXTRACCIÓN .....	64
5.2.2	ELIMINACIÓN DE PRONOMBRES .....	64
5.2.3	ELIMINACIÓN DE DEÍCTICOS .....	65
5.2.4	ELIMINACIÓN DE FORMAS IMPERSONALES O INCOMPLETAS .....	65
5.2.5	ELIMINACIÓN DE OCURRENCIAS REPETIDAS.....	66
5.2.6	ELIMINACIÓN DE CIFRAS .....	66
5.2.7	ELIMINACIÓN DE NOMBRES PROPIOS .....	67
5.2.8	ELIMINACIÓN DE LOS VERBOS MODALES.....	67
5.3.	ETAPA 3: ETIQUETADO DE LAS OCURRENCIAS DEPURADAS.....	68
5.3.1.	ANÁLISIS SINTÁCTICO .....	69
5.3.2.	ETIQUETAJE SEMÁNTICO .....	69
5.3.2.1.	TIPO DE VERBO .....	69
5.3.2.2.	SENTIDO DE VERBO .....	70
5.3.2.3.	ÁREA TEMÁTICA .....	70
5.4.	ASIGNACIÓN DE LAS CARGAS DE POLARIDAD.....	71
<b>PARTE IV: ANÁLISIS DE DATOS Y RESULTADOS .....</b>		<b>72</b>

CAPÍTULO 6: ANÁLISIS DE DATOS Y RESULTADOS .....	73
6.1.    ANÁLISIS DE DATOS Y RESULTADOS .....	74
6.1.1.    DEL UNIVERSO A LA MUESTRA DE TRABAJO.....	74
6.1.2.    DISTRIBUCIÓN DE LAS OCURRENCIAS POR PERIÓDICO .....	76
6.1.3.    FRECUENCIA DE LOS VERBOS .....	78
6.1.4.    FRECUENCIA SEGÚN EL TIPO DE VERBO .....	80
6.1.4.1.    VERBOS FRASEOLÓGICOS.....	80
6.1.4.2.    VERBOS CONECTORES .....	81
6.1.4.3.    VERBOS DISCURSIVOS .....	82
6.1.5.    CARGAS SEGÚN EL TIPO DEL VERBO.....	83
6.1.5.1.    POSITIVO .....	85
6.1.5.2.    NEGATIVO .....	85
6.1.5.3.    NEUTRO .....	85
6.1.6.    FRECUENCIA SEGÚN EL SENTIDO DEL VERBO.....	86
6.1.6.1. <i>POSSESSION</i> .....	89
6.1.6.2. <i>COMMUNICATION</i> .....	90
6.1.6.3. <i>CHANGE</i> .....	91
6.1.6.4. <i>SOCIAL</i> .....	93
6.1.6.5. <i>CREATION</i> .....	94
6.1.6.6. <i>PERCEPTION</i> .....	95
6.1.6.7. <i>STATIVE</i> .....	96
6.1.6.8. <i>COGNITION</i> .....	97
6.1.6.9. <i>MOTION</i> .....	99
6.1.7.    CARGAS DE LOS VERBOS .....	100
6.1.7.1.    VERBOS POSITIVOS.....	101
6.1.7.2.    VERBOS NEGATIVOS .....	102
6.1.7.3.    VERBOS NEUTROS.....	103
6.1.8.    CARGAS DEL OBJETO DIRECTO .....	104
6.1.9.    ÁREA TEMÁTICA DE LOS OBJETOS DIRECTOS .....	106

6.1.9.1.	ÁREA TEMÁTICA <i>ECONOMÍA</i> SEGÚN DISTRIBUCIÓN DE LAS CARGAS DE LOS OBJETOS DIRECTOS .....	107
6.1.9.2.	ÁREA TEMÁTICA <i>SOCIEDAD</i> SEGÚN DISTRIBUCIÓN DE LAS CARGAS DE LOS OBJETOS DIRECTOS .....	109
6.1.9.3.	ÁREA TEMÁTICA <i>POLÍTICA</i> SEGÚN DISTRIBUCIÓN DE LAS CARGAS DE LOS OBJETOS DIRECTOS .....	110
6.1.9.4.	ÁREA TEMÁTICA <i>GENERAL</i> SEGÚN DISTRIBUCIÓN DE LAS CARGAS DE LOS OBJETOS DIRECTOS .....	112
6.1.10.	CARGAS OBJETOS DIRECTOS CON RELACIÓN AL ÁREA TEMÁTICA .....	113
6.1.10.1.	OBJETOS DIRECTOS POSITIVOS CON RELACIÓN AL ÁREA TEMÁTICA .....	114
6.1.10.2.	OBJETOS DIRECTOS NEGATIVOS CON RELACIÓN AL ÁREA TEMÁTICA .....	115
6.1.10.3.	OBJETOS DIRECTOS NEUTROS CON RELACIÓN AL ÁREA TEMÁTICA .....	116
6.1.11.	CARGA TOTAL PATRÓN [VERBO + OBJETO DIRECTO].....	116
6.1.12.	CARGA TOTAL PATRÓN [VERBO + OBJETO DIRECTO] CON RELACIÓN AL TIPO DEL VERBO	121
6.1.12.1.	POSITIVO.....	121
6.1.12.2.	NEGATIVO .....	122
6.1.12.3.	NEUTRO.....	122
6.1.13.	CARGA TOTAL PATRÓN [VERBO + OBJETO DIRECTO] CON RELACIÓN AL SENTIDO DEL VERBO	123
6.1.13.1.	POSITIVO .....	123
6.1.13.2.	NEGATIVO .....	125
6.1.13.3.	NEUTRO .....	126
6.1.14.	CARGA TOTAL PATRÓN [VERBO + OBJETO DIRECTO] CON RELACIÓN AL ÁREA TEMÁTICA .....	127
6.1.14.1.	POSITIVO .....	127
6.1.14.2.	NEGATIVO .....	128
6.1.14.3.	NEUTRO .....	130

6.1.15. RECAPITULACIÓN .....	131
<b>PARTE V: CONCLUSIONES .....</b>	<b>133</b>
CAPÍTULO 7: CONCLUSIONES .....	134
7.1. CONCLUSIONES .....	134
7.1.1. VALIDACIÓN DE LOS OBJETIVOS .....	135
7.1.2. VALIDACIÓN DE LAS HIPÓTESIS .....	138
7.2. APORTE DE LA INVESTIGACIÓN AL CAMPO DE ESTUDIO .....	139
7.3. LIMITACIONES Y PERSPECTIVAS DE TRABAJO .....	140
BIBLIOGRAFÍA.....	141
ANEXOS .....	150
ANEXO A: FRECUENCIA DE VERBOS EN LA MUESTRA.....	150
ANEXO B: REGLAS DE COMBINACIÓN SEGÚN LAS CARGAS POSITIVAS, NEGATIVAS O NEUTRAS.....	153
ANEXO C: REGLAS DE COMBINACIÓN EN LOS DATOS CON EJEMPLOS .....	154

## LISTA DE FIGURAS

<i>Figura 1: Tipos de oración en Gili Gaya</i> .....	27
<i>Figura 2: Relaciones verbales en WordNet</i> .....	31
<i>Figura 3: Tipos de objetos directos Dapena (2002)</i> .....	38
<i>Figura 4: Distribución de artículos por periódico</i> .....	52
<i>Figura 5: Esquema metodología de trabajo</i> .....	57
<i>Figura 6: Nombres campos base de datos</i> .....	58
<i>Figura 7: Distribución tipos de verbos</i> .....	61
<i>Figura 8: Interfaz WordNet 2.1</i> .....	62
<i>Figura 9: Ejemplo árbol sintáctico</i> .....	68
<i>Figura 10: Comportamiento ocurrencias del patrón [verbo + objeto directo]</i> .....	76
<i>Figura 11: Distribución de ocurrencias por periódico</i> .....	76
<i>Figura 12: Frecuencia verbos por ocurrencias</i> .....	79
<i>Figura 13: Verbos fraseológicos</i> .....	81
<i>Figura 14: Verbos conectores</i> .....	82
<i>Figura 15: Verbos discursivos</i> .....	83
<i>Figura 16: Frecuencia sentido de los verbos</i> .....	87
<i>Figura 17: Dimensiones del IMP en Colombia</i> .....	88
<i>Figura 18: Frecuencia verbos de sentido possession</i> .....	89
<i>Figura 19: Frecuencia verbos de sentido communication</i> .....	91
<i>Figura 20: Frecuencia verbos de sentido change</i> .....	92
<i>Figura 21: Frecuencia verbos de sentido social</i> .....	93
<i>Figura 22: Frecuencia verbos de sentido creation</i> .....	94
<i>Figura 23: Frecuencia verbos de sentido perception</i> .....	96
<i>Figura 24: Frecuencia verbos de sentido stative</i> .....	97
<i>Figura 25: Frecuencia verbos de sentido cognition</i> .....	98
<i>Figura 26: Frecuencia verbos de sentido motion</i> .....	100
<i>Figura 27: Frecuencia de las ocurrencias según carga del verbo</i> .....	101
<i>Figura 28: Frecuencia de ocurrencias con verbos positivos</i> .....	102

<i>Figura 29: Frecuencia de ocurrencias con verbos negativos .....</i>	<i>103</i>
<i>Figura 30: Frecuencia de ocurrencias con verbos neutros .....</i>	<i>104</i>
<i>Figura 31: Distribución objetos directos según carga asignada .....</i>	<i>105</i>
<i>Figura 32: Distribución frecuencia de las áreas temáticas .....</i>	<i>107</i>
<i>Figura 33: Área temática economía según las cargas de los objetos directos.....</i>	<i>108</i>
<i>Figura 34: Área temática sociedad según la carga de los objetos directos.....</i>	<i>109</i>
<i>Figura 35: Área temática política según cargas de los objetos directos .....</i>	<i>111</i>
<i>Figura 36: Área temática general según cargas de los objetos directos .....</i>	<i>112</i>
<i>Figura 37: Objetos directos positivos por área temática.....</i>	<i>114</i>
<i>Figura 38: Objetos directos negativos por área temática.....</i>	<i>115</i>
<i>Figura 39: Objetos directos neutros por área temática .....</i>	<i>116</i>
<i>Figura 40: Carga final de las ocurrencias del patrón con premodificación negativa.....</i>	<i>118</i>
<i>Figura 41: Frecuencia carga final ocurrencias.....</i>	<i>119</i>
<i>Figura 42: Frecuencia reglas de combinación según ocurrencias .....</i>	<i>120</i>
<i>Figura 43: Área temática del objeto directo según carga positiva del patrón.....</i>	<i>127</i>
<i>Figura 44: Área temática del objeto directo según carga negativa de la ocurrencia.....</i>	<i>129</i>
<i>Figura 45: Área temática del objeto directo según carga neutra del patrón.....</i>	<i>130</i>

## LISTA DE TABLAS

<i>Tabla 1: Número de palabras, synsets y sentidos WordNet 2.1</i> .....	30
<i>Tabla 2: Categorías de clasificación verbal en WordNet</i> .....	31
<i>Tabla 3: Cantidad de artículos recuperados por periódico</i> .....	50
<i>Tabla 4: Cantidad de artículos obtenidos en la muestra</i> .....	51
<i>Tabla 5: Número de artículos y muestras de los periódicos</i> .....	52
<i>Tabla 6: Ejemplos de ocurrencias eliminadas: errores de extracción</i> .....	64
<i>Tabla 7: Ejemplos de ocurrencias eliminadas: pronombres</i> .....	64
<i>Tabla 8: Ejemplos de ocurrencias eliminadas: deícticos</i> .....	65
<i>Tabla 9: Ejemplos de ocurrencias eliminadas: formas impersonales o incompletas</i> .....	65
<i>Tabla 10: Ejemplos de ocurrencias repetidas eliminadas</i> .....	66
<i>Tabla 11: Ejemplos de ocurrencias con cifras</i> .....	66
<i>Tabla 12: Reglas con ocurrencias que contienen cifras</i> .....	66
<i>Tabla 13: Ejemplos de ocurrencias con nombres propios</i> .....	67
<i>Tabla 14: Ejemplo ocurrencias extraídas</i> .....	69
<i>Tabla 15: Ejemplos etiquetaje del tipo de verbo</i> .....	69
<i>Tabla 16: Ejemplos etiquetaje del sentido de verbo</i> .....	70
<i>Tabla 17: Ejemplos etiquetaje del área temática del objeto directo</i> .....	70
<i>Tabla 18: Ejemplos etiquetaje de la polaridad de los elementos de las ocurrencias</i> .....	71
<i>Tabla 19: del universo de datos a la muestra de trabajo</i> .....	75
<i>Tabla 20: Distribución cargas patrón por periódico</i> .....	77
<i>Tabla 21: Frecuencia de las ocurrencias con relación al tipo de verbo</i> .....	80
<i>Tabla 22: Frecuencia cargas ocurrencias según tipo del verbo</i> .....	84
<i>Tabla 23: Verbos fraseológicos según carga de polaridad</i> .....	85
<i>Tabla 24: Verbos conectores según carga de polaridad</i> .....	85
<i>Tabla 25: Verbos conectores según carga de polaridad</i> .....	85
<i>Tabla 26: Frecuencia sentidos verbo dar</i> .....	99
<i>Tabla 27: Distribución objetos directos según área temática</i> .....	113
<i>Tabla 28: Tipo de verbo según carga positiva del patrón</i> .....	121

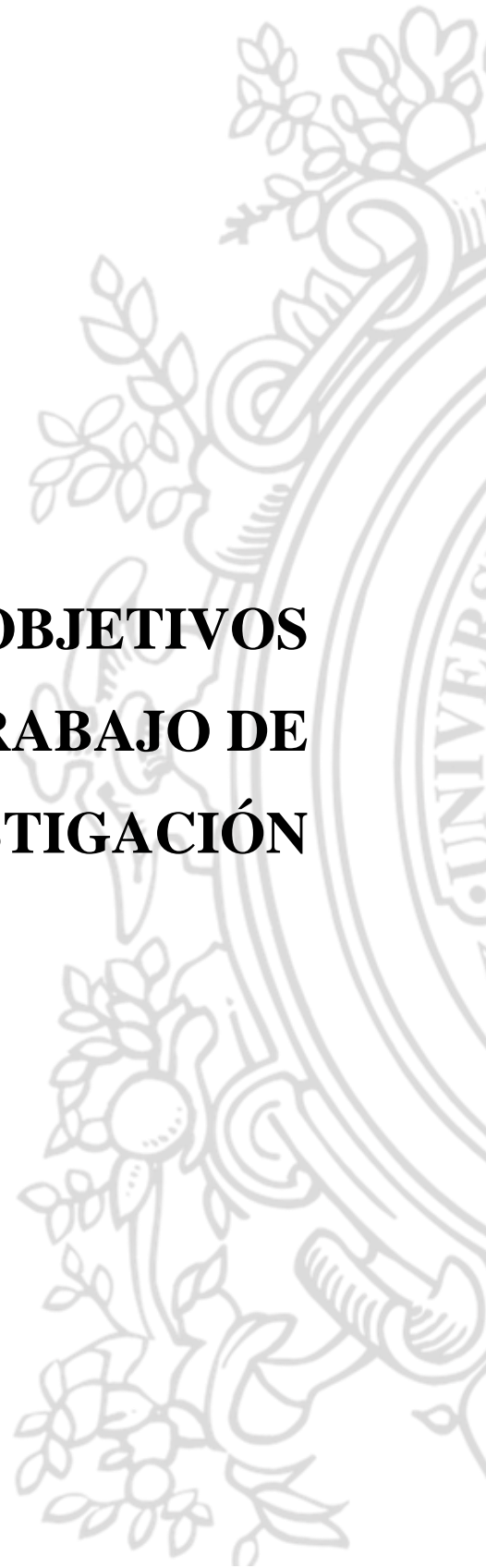


<i>Tabla 29: Tipo de verbo según carga negativa del patrón.....</i>	<i>122</i>
<i>Tabla 30: Tipo de verbo según carga neutra del patrón.....</i>	<i>122</i>
<i>Tabla 31: Sentido del verbo según carga positiva del patrón.....</i>	<i>124</i>
<i>Tabla 32: Sentido del verbo según carga negativa del patrón.....</i>	<i>125</i>
<i>Tabla 33: Sentido del verbo según carga neutra del patrón.....</i>	<i>126</i>

## LISTA DE ANEXOS

<i>Anexo A: Frecuencia de verbos en la muestra .....</i>	<i>150</i>
<i>Anexo B: Reglas de combinación según las cargas positivas, negativas o neutras .....</i>	<i>153</i>
<i>Anexo C: Reglas de combinación en los datos con ejemplos .....</i>	<i>154</i>

**PARTE I: INTRODUCCIÓN, OBJETIVOS  
Y ESBOZO DEL TRABAJO DE  
INVESTIGACIÓN**



# CAPÍTULO 1: INTRODUCCIÓN

“The Analytical Engine has no pretensions whatever to originate anything. **It can do whatever we know how to order it to perform.** It can follow analysis; but it has no power of anticipating any analytical relations or truths. Its province is to assist us to making available what we are already acquainted with”. *Ada King Lovelace*

Tomar decisiones a la hora de adquirir productos o servicios se constituye en una de las actividades humanas que involucra, en gran medida, el punto de vista de los demás. Las recomendaciones, sugerencias o experiencias de compra constituyen la base de lo que podemos denominar “lo que otros piensan” (Pang & Lee, 2008, p. 1). En el pasado, la oralidad era el medio prevalente al momento de solicitar opiniones sobre algún producto o servicio que se deseaba adquirir. No obstante, en la actualidad, las nuevas tecnologías de la información desempeñan un rol determinante en este proceso.

En este orden de ideas, en las últimas décadas, las tendencias de compra han dado un giro significativo. Por ejemplo, el continuo aumento en el uso de la Internet aplicado al *Marketing* ha dado origen al denominado *e-commerce* o comercio virtual (McLuhan & Fiore, 1988), generando un crecimiento considerable de compras *online* como una alternativa cada vez más utilizada por los consumidores, en constante desarrollo por parte de los vendedores y que se caracteriza por: la creación de blogs, foros y reseñas de productos.

Todo esto dio origen a un flujo considerable de información y como consecuencia, a la producción de grandes cantidades de datos en bruto a la espera de ser procesados desde distintos enfoques, tanto teóricos como metodológicos.

Recientemente, investigadores han explorado varias alternativas de solución al desafío que presenta el analizar la totalidad de reseñas, comentarios u opiniones sobre personas, eventos, organizaciones, productos o temas. Este problema ha sido denominado *Opinion Mining* (Minería de Opinión) o *Sentiment Analysis* (Análisis de Sentimientos/Análisis de Opinión) (Ding, Liu, & Yu, 2008) y ha sido abordado principalmente por las ciencias del lenguaje y la computación.

Según Liu (2012), el Análisis de Sentimientos (en adelante AS), también denominado minería de opinión, es el campo de estudio que analiza las opiniones de la gente, sus sentimientos, evaluaciones, apreciaciones, actitudes y emociones sobre productos, servicios, organizaciones, individuos, asuntos, eventos, temas y atributos (p.1).

Tanto el término AS como Minería de Opinión son utilizados sin distinción en la literatura académica. En el caso de la industria solo aplica el primero (Liu, 2012, p. 1). Como consecuencia, para los propósitos del presente trabajo de investigación, se empleará Análisis de Sentimientos.

El AS se ha enfocado en la clasificación de textos de forma automática, principalmente, en reseñas de productos o servicios; sin dejar de lado otras aplicaciones en ámbitos políticos, militares, económicos y/o sociales, con el objetivo de etiquetar cada texto como positivo, negativo o neutro. Esta información ha sido clave en la mejora de la oferta por parte de las empresas quienes buscan vender sus productos o servicios que cumplan con las expectativas de los compradores potenciales o definir las preferencias de las personas en los ámbitos arriba mencionados. De este modo, desde el nivel pragmático de la lengua, el AS se entiende como la clasificación automática de textos según la carga de opinión expresada en el mismo.

## **1.1. HIPÓTESIS**

En la mayoría de los trabajos consultados sobre AS, el foco de atención se ha dirigido a los adjetivos como unidades lingüísticas poseedoras de opinión, relegando a un segundo plano a las demás partes del habla (POS, por sus siglas en inglés). Sin embargo, en la actualidad, ya se han empezado a realizar diversas investigaciones, teniendo en cuenta las diferentes categorías gramaticales, hasta ahora poco trabajadas.

Con relación a nuestro trabajo de investigación, presentamos las siguientes hipótesis:

En primer lugar, consideramos que el verbo es, al igual que el adjetivo, un elemento poseedor de carga de opinión, merecedor de un tratamiento especial en la construcción de las herramientas para el AS. Por lo tanto, hemos definido el patrón [verbo + objeto directo] como unidad de análisis de nuestro trabajo de investigación.

En segundo lugar, pensamos que al tener en cuenta solamente la información aportada por las ocurrencias del patrón [verbo + objeto directo], es posible identificar la forma como se aborda el fenómeno de la pobreza en la prensa colombiana.

## **1.2. OBJETIVOS**

El presente trabajo de investigación, en el ámbito de la lingüística computacional, se centra en el aspecto lingüístico del procesamiento de lenguaje natural. Para ello, se ha planteado el siguiente objetivo general y tres específicos.

### **1.2.1. GENERAL**

Diseñar una metodología para la construcción de una base de datos con patrones [verbo + objeto directo], que permita la discriminación de estos como unidad lingüística para el mejoramiento de la clasificación automática de artículos de prensa, por medio de las técnicas de Análisis de Sentimientos.

### **1.2.2. ESPECÍFICOS**

- ❖ Caracterizar las ocurrencias del patrón [verbo + objeto directo] para propósitos de análisis de sentimientos, identificando sus rasgos distintivos y discriminantes.
- ❖ Definir una metodología de depuración de las ocurrencias del patrón [verbo + objeto directo] de tal modo que puedan extraerse los elementos más discriminantes en la construcción de las herramientas para Análisis de Sentimientos.
- ❖ Describir cómo es abordado el tema de la pobreza en la prensa colombiana mediante la descripción lingüística de las ocurrencias del patrón [verbo + objeto directo] para propósitos de Análisis de Sentimientos.

### **1.3. ESBOZO DEL TRABAJO DE INVESTIGACIÓN**

Esta investigación se divide en cuatro partes, las cuales se describen a continuación.

La Parte I: se presenta una breve introducción al tema de Análisis de Sentimientos al igual que la hipótesis de nuestro trabajo y los objetivos de investigación.

La Parte II: marco teórico, está dividido en dos capítulos: análisis de sentimientos y el verbo en español. Allí se describen las nociones más importantes sobre el análisis de sentimientos y las características del verbo con propósitos de análisis de sentimientos.

La parte III: metodología de investigación, contiene 2 capítulos. En el primero, se presenta el paso a paso seguido en la edificación del presente trabajo de investigación. En el segundo, se presenta la propuesta de guía metodológica en la construcción de una base de datos [verbo + objeto directo] que puede ser utilizada en las aplicaciones de análisis de sentimientos.

La parte IV: análisis de datos y resultados. Se describen los hallazgos al analizar los datos siguiendo la propuesta de identificación y depuración de las ocurrencias del patrón [verbo + objeto directo]. Asimismo, se presentan las precisiones producto del análisis de los datos y se concluye con una recapitulación sintetizando los datos obtenidos de acuerdo a las hipótesis y objetivos del presente trabajo.

La parte V: conclusiones. Se presentan las consideraciones finales sobre el trabajo de investigación junto a las limitaciones y las perspectivas de trabajo.

**PARTE II: ANÁLISIS DE SENTIMIENTOS  
Y EL VERBO EN ESPAÑOL**





## CAPÍTULO 2: ANÁLISIS DE SENTIMIENTOS

En este capítulo, abordaremos el AS desde la definición, su objeto de estudio y concluyendo con los diversos enfoques y métodos de trabajo empleados en la solución de los desafíos propios del campo vistos desde las distintas áreas del conocimiento implicadas.

### 2.1. ¿QUÉ ES EL ANÁLISIS DE SENTIMIENTOS?

Si bien las investigaciones en AS datan de principios del siglo XXI en los trabajos de (J.M. Wiebe, 2000), (Das & Chen, 2007), (Morinaga, Yamanishi, Tateishi, & Fukushima, 2002), (Pang, Lee, & Vaithyanathan, 2002), (Turney, 2002), entre otros; y que se registren trabajos relacionados al campo en la década de 1990 como la interpretación de metáforas (Janyce M Wiebe, 1990) (J. Wiebe, 1994), extracción de adjetivos de sentimiento (Hearst, 1992), computación afectiva (Vasileios Hatzivassiloglou & McKeown, 1997) y análisis de subjetividad, puntos de vista y afectos (Picard, 2003) (J. M. Wiebe, Bruce, & O'Hara, 1999); solo hasta el año 2003 se comienza a registrar el término Análisis de Sentimientos en los trabajos de (Nasukawa & Yi, 2003) al igual que el término Minería de Opinión en los escritos de (Dave, Lawrence, & Pennock, 2003).

Para Liu (2015), el Análisis de Sentimientos o Minería de Opinión es el campo de estudio que analiza las opiniones de la gente, sus sentimientos, evaluaciones, apreciaciones, actitudes, y emociones sobre productos, servicios, organizaciones, individuos, asuntos, eventos, temas y atributos (p.1).

En la actualidad, hay un amplio abanico de posibilidades para referirse a esta tarea: Análisis de Sentimientos, Minería de Opinión, Análisis de Opinión, Extracción de Opinión, Minería de Sentimientos, Análisis de Subjetividad, Análisis de Afecto, Análisis de Emoción, Minería de Reseñas; es recogido bajo el término Análisis de Sentimientos.

En este punto se presenta una primera dificultad al establecer las diferencias entre sentimiento y opinión, dado a que la denominación de AS procede más desde el ámbito de las ciencias de la computación que desde la lingüística; de allí que se origine la amplia gama de denominaciones previamente listadas (Liu, 2015, p. 2). Según el autor, la diferencia es bastante sutil: mientras la opinión corresponde a la forma concreta como una persona percibe algo, el sentimiento es más

una implicación emocional. Es decir, mientras la oración “la operación de Colpensiones ha sido un desastre” expresa una opinión; el enunciado “Sin embargo, se va preocupado por la falta de garantías para los sectores más vulnerables y la supresión de derechos por el conflicto” expresa un sentimiento. Por ende, el AS alude a la clasificación automática de textos en positivos, negativos o neutros.

## **2.2. NIVELES DE ANÁLISIS**

Respecto a los principales niveles en los cuales se ha estudiado el AS, estos pueden ser observados desde lo general a lo particular; es decir, la clasificación puede darse a nivel de documento, a nivel de oración o a nivel entidad-aspecto. A continuación, se presenta una descripción de cada uno de los niveles trabajados desde el análisis de sentimientos.

### **2.2.1. ANÁLISIS DE SENTIMIENTOS A NIVEL DE DOCUMENTO**

Este primer nivel de análisis, también denominado clasificación de sentimiento a nivel de documento, responde a la tarea de identificar si un documento de opinión completo expresa un sentimiento positivo o negativo (Pang *et al.*, 2002) y (Turney, 2002).

En este nivel de clasificación, se asume implícitamente que el documento expresa opiniones con relación a una sola entidad (tema) y de allí sus limitaciones, cuando los textos a clasificar son comparaciones o implican distintas entidades y se hace necesario recurrir a los otros dos niveles de análisis, cuyo enfoque es mucho más minucioso.

### **2.2.2. ANÁLISIS DE SENTIMIENTOS A NIVEL DE ORACIÓN**

El segundo nivel de análisis busca clasificar cada oración en positivo, negativo o neutro; entendiendo este último como “no opinión” (Liu, 2015, p.9). En la clasificación de subjetividad de (J. M. Wiebe *et al.*, 1999), una aproximación a este tipo de análisis, se distinguen dos tipos de oración: objetivas y subjetivas. Sin embargo, subjetividad no es equivalente a opinión o sentimiento, esto en concordancia con la definición de opinión y sentimiento presentada previamente (§ 2.1.).

Este nivel de análisis presenta mayor dificultad debido a que la cantidad de información para analizar es menor que la presentada a nivel de documento. Del mismo modo, en este nivel de análisis, la categoría neutra es omitida con frecuencia debido a que presenta un mayor desafío en la clasificación de polaridad triclase (positivo, negativo, neutro). Sin embargo, a este nivel, lo neutro presenta mayor relevancia, debido a la gran cantidad de oraciones que no expresan opinión ni positiva ni negativa, siendo este tipo de oraciones, carentes de opinión, las que se denominan neutras (Liu, 2015, p. 70).

En nuestro trabajo abordamos el caso del neutro como un componente importante, ofreciendo una propuesta de descripción de los elementos textuales que consideramos como neutros. De igual forma, proporcionamos ejemplos, extraídos desde nuestros datos, para ilustrar cada uno de los aspectos marcados con esta polaridad.

Otro factor importante en este nivel de análisis, consiste en que este no está en directa relación con la entidad de la cual se originan las opiniones en el texto. Aunque es un modelo viable, presenta falencias, puesto que se corre el riesgo de tomar el sentimiento de una oración y asignarlo a otra cuya entidad es distinta a la inicial.

El tercer modelo presenta una propuesta más detallada al buscar el establecimiento de la relación entidad-aspecto, recurriendo de igual manera a técnicas de Procesamiento de Lenguaje Natural (PLN) avanzadas, para la extracción de los elementos de análisis.

### **2.2.3. ANÁLISIS DE SENTIMIENTOS A NIVEL DE ENTIDAD-ASPECTO**

Denominado en principio Análisis de Características (*Feature Level*) en los trabajos de (Hu & Liu, 2004) y (Liu, 2010); este nivel de análisis busca directamente la opinión y el receptor (*target*) de la misma (Liu, 2015, p. 91).

En este nivel, se refuta la idea de que si una oración es factible de ser caracterizada como positiva, negativa o neutra, todos los elementos en ella serían en principio marcados con la misma polaridad. Sin embargo, este razonamiento no aplica, dado que una oración puede contener múltiples opiniones, como se observa en el fragmento de la columna de economía del periódico El Espectador (*La meta: pensionar más colombianos*) “Será el principio de una apuesta para mejorar

la calidad de un modelo que a lo largo de estos últimos años ha dejado en evidencia enormes falencias, el causante de que un porcentaje mayoritario de la población esté sin pensión”.

### **2.3. APLICACIONES DEL ANÁLISIS DE SENTIMIENTOS**

Aunque es un área de reciente explotación tanto en la investigación como en el sector industrial, el AS cuenta hoy en día con diversas aplicaciones en ámbitos incluso gubernamentales.

Es factible establecer el punto de quiebre que potenció la reciente explotación de los alcances del AS. El antes y el después del año 2000, marca el flujo de información susceptible de ser analizada. Antes, los textos de opinión no eran guardados de forma digital. Ahora, con la expansión de la Internet y las Redes Sociales, se cuenta con un flujo constante de información almacenada digitalmente (Liu, 2015, p. 3), lo que ha facilitado su fácil acceso y permitiendo el procesamiento de los datos.

En el ámbito gubernamental, las aplicaciones se enfocan más en identificar la favorabilidad en las políticas de gobierno con el ánimo de conservar lo positivo y mejorar los elementos necesarios. Así mismo, en épocas electorales la utilización del análisis de sentimientos se convierte en una herramienta que permite rastrear las tendencias electoras de los votantes y predecir el o los candidatos a ganar (Liu, 2015, p. 4).

En el ámbito empresarial, el AS se convierte en una importante herramienta para conocer la percepción de las personas frente al uso o consumo de los productos o servicios ofrecidos, con miras a potenciar la oferta y mejorar continuamente la calidad de los negocios. Del mismo modo, un constante monitoreo de la opinión de los usuarios permite a las empresas generar diagnósticos de posicionamiento de sus productos y tener un punto de referencia frente a la competencia.

A nivel de los individuos, el AS permite obtener una amplia cantidad de información, local e internacional, con relación a un producto o servicio que se quiere adquirir. Estos datos pueden ser del tipo experiencias de compra, fortalezas y debilidades sobre el producto o servicio, favoreciendo un conocimiento compartido sobre el mismo.

En síntesis, las aplicaciones del AS pueden abarcar desde el comercio hasta asuntos gubernamentales, pasando por el campo de la salud, el turismo o los servicios financieros. Análisis

de reseñas, predicción del éxito o fracaso de productos o servicios, inversiones en la Bolsa de valores, comportamiento de las personas en épocas electorales, son algunos de los usos en los ámbitos mencionados anteriormente. Inclusive, cientos de empresas han construido o están en proceso de construir sus propios programas para Análisis de Sentimientos: Google, Microsoft, Hewlett-Packard, son algunos ejemplos (Liu, 2015, p. 6).

## **2.4. EL ANÁLISIS DE SENTIMIENTOS DESDE DIFERENTES PERSPECTIVAS**

Las ciencias de la computación y la lingüística son las dos áreas del conocimiento que confluyen principalmente en la construcción de las distintas metodologías para solucionar los problemas inherentes al AS dando origen a distintos enfoques de trabajo.

Pang & Lee (2008) presentan una recopilación detallada de los diversos enfoques empleados para dar solución a los desafíos del AS. En primer lugar, se presentan los componentes lingüísticos tenidos en cuenta para la construcción de los vectores de características que representan un texto. En segundo lugar, se listan los enfoques más importantes desde las ciencias de la computación para dar solución al problema de SA.

### **2.4.1. EL ANÁLISIS DE SENTIMIENTOS DESDE LA PERSPECTIVA LINGÜÍSTICA**

La lingüística se convierte en la base desde la cual se construyen los distintos patrones de análisis, a partir de los cuales se representan los textos como vectores de características y se logra la clasificación según la polaridad (positivo, negativo, neutro). En términos computacionales, lo que se pretende con el AS es resolver un problema de clasificación, generalmente a través de modelos de Aprendizaje Automático Supervisado.

#### **2.4.1.1. PRESENCIA DE TÉRMINO VS. FRECUENCIA**

En términos generales, lo que se pretende con el AS es clasificar un texto de acuerdo a su carga de opinión. Un primer enfoque lingüístico consiste en la identificación de los términos presentes en el texto (Pang *et al.*, 2002), a través del cual se construyen los patrones de análisis que permiten construir los vectores de características que representan cada texto, utilizando

medidas de peso conocidas como *term frequency – inverse document frequency* (Baeza-Yates & Ribeiro-Neto, 1999).

La diferencia entre el análisis por presencia de término y el análisis por frecuencia es que, en el primero, un solo término puede ser determinante en la polaridad del texto; mientras que, en el segundo, la frecuencia relativa de aparición de las palabras define la temática del texto (Pang *et al.*, 2002).

#### **2.4.1.2. CARACTERÍSTICAS BASADAS EN TÉRMINOS MÁS ALLÁ DE LOS UNIGRAMAS**

Al hablar de n-gramas (patrones sintácticos con n tokens - unidades léxicas-) es claro que nos movemos entre el léxico compuesto y la sintaxis. Entendemos por *token* “*any instance of a particular wordform in a text*” opuesto al concepto de *type*, el cual es definido como “*a particular, unique wordform*” (McEnery & Wilson, 2001, p. 50).

Distintas aplicaciones del AS convergen alrededor del léxico, dependiendo del objetivo perseguido. Al hablar de reseñas de películas Pang *et al.* (2002) reportan mayor efectividad de los unigramas en contraste con los bigramas. Sin embargo, Dave *et al.* (2003), con relación a la clasificación de revisiones de productos, expresan que los bigramas y trigramas son mejores características en este punto.

Otro aspecto importante es la posición al interior del texto (al principio, en medio o al final) de las unidades léxicas. En (Pang & Lee, 2008) y (Kim & Hovy, 2006), la información aportada por la posición del token es tenida en cuenta en la construcción de los vectores de características.

La literatura reporta otras técnicas basadas en los n-gramas: por un lado, definiendo las características de las unidades léxicas y estableciendo las relaciones entre ellas (Riloff, Patwardhan, & Wiebe, 2006). Por otro lado, construyendo una estructura de dependencia en la codificación de vocabulario para las polaridades positiva o negativa (M. Airoidi, Bai, Padman, & Airoidi, 2006). Finalmente, utilizando la distancia contrastiva entre términos para definir las características para la clasificación automática (Snyder & Barzilay, 2007).

### **2.4.1.3. INFORMACIÓN DE LAS CATEGORÍAS GRAMATICALES (*PART-OF-SPEECH INFORMATION, POS*)**

La información aportada por las categorías gramaticales es parte fundamental en la solución de los desafíos en el AS, debido al potencial en la desambiguación del sentido de las palabras (aspecto clave al realizar los análisis), lo que permite pasar de un nivel léxico y sintáctico a uno de carácter semántico.

Los primeros trabajos utilizando las categorías gramaticales (Hatzivassiloglou & McKeown, 1997) toman al adjetivo como unidad de análisis. De hecho, esta unidad gramatical ha sido de gran explotación en el área por parte de varios investigadores (Mullen & Collier, 2004), (Whitelaw, Garg, & Argamon, 2005). Trabajos posteriores en la detección de subjetividad mostraron una importante correlación entre el uso de adjetivos y las frases subjetivas (Hatzivassiloglou & Wiebe, 2000).

En síntesis, los adjetivos se han configurado como un elemento clave en la detección de la opinión. Inclusive, hay trabajos en los cuales no sólo se utiliza el adjetivo, sino que además estos son vistos en unión con otras categorías gramaticales como el adverbio (Turney, 2002).

Por último, a pesar de la evidente efectividad del adjetivo en la detección de la opinión, no se han descartado otras categorías gramaticales de las cuales ya existen trabajos con resultados importantes. Es el caso de los sustantivos subjetivos en (Riloff, Wiebe, & Wilson, 2003) o de los adverbios y de los verbos, además de otras categorías en (Riloff et al., 2003), (Benamara, Cesarano, Picariello, Reforgiato, & Subrahmanian, 2007) y (Nasukawa & Yi, 2003).

El objeto de nuestra investigación, apunta a caracterizar las ocurrencias del patrón [verbo + objeto directo] para la construcción de bases de datos con propósitos de Análisis de Sentimientos, describiendo Los elementos más relevantes a tener presentes en la construcción de vectores de características discriminantes para dicho propósito.

### **2.4.1.4. SINTAXIS**

Las relaciones sintácticas, más allá de mejorar los resultados con relación a las técnicas basadas en la utilización de los n-gramas, ha permitido combinar ambas técnicas y aportar a la

solución de los problemas de clasificación de documentos. No obstante, son opuestos los resultados obtenidos con relación a la efectividad en cuanto a la información de dependencia (Dave et al., 2003), (Gamon, 2004), (Matsumoto, Takamura, & Okumura, 2005) y (Ng, Dasgupta, & Arifin, 2006).

Otro aporte esencial de la sintaxis al campo del AS, ha sido la posibilidad de identificar y modelar aspectos como la negación, los intensificadores y los términos que indican disminución (Kennedy & Inkpen, 2006). Patrones sintácticos complejos al igual que colocaciones han sido identificados como elementos relevantes en la detección de la subjetividad en los textos (Nasukawa & Yi, 2003) y (Riloff & Wiebe, 2003).

### ❖ LA NEGACIÓN

Al hablar de la negación, es necesario recurrir a un aspecto que ha demostrado ser vital en el análisis de sentimientos, a pesar de ser problemático. Desde varios enfoques como el de *bag-of-words*, las oraciones del tipo “no crear impuesto” y “crear impuesto” son consideradas similares en cuanto solo difieren en un token. Sin embargo, la negación obliga a que ambas oraciones sean marcadas con dos cargas distintas.

Al respecto, (Das & Chen, 2007) agregan el adverbio “no” a las palabras cercanas a una negación. No obstante, esta regla no siempre invierte la polaridad de la oración, lo que genera un problema en la aplicación de este modelo. Por otro lado, en (Na, Sui, Khoo, Chan, & Zhou, 2004) se registra una mejora en la eficiencia de la clasificación al integrar patrones morfosintácticos específicos para marcar las oraciones completas como negativas, según los diferentes términos para negar.

Otro aspecto importante en la solución de los problemas al abordar la negación, hace referencia al uso sutil de las estrategias discursivas en la construcción de los textos. En este punto, la identificación de los sarcasmos e ironías constituye un desafío interesante. Inclusive, frente a la ausencia de estas estrategias retóricas, hay otros casos más difíciles de solucionar. (Hoffmann, Wilson, & Wiebe, 2009) abordan el tema de la negación argumentando que esta no siempre es directa. En efecto, los autores señalan que la negación puede ser local, implicar dependencias de mayor distancia o intensificar en vez de cambiar la polaridad.



#### **2.4.1.5. CARACTERÍSTICAS ORIENTADAS A TEMAS (*TOPIC-ORIENTED FEATURES*)**

Las técnicas de AS basadas en la interacción entre un tema y la opinión (*topic-sentiment interaction*) apuntan a establecer la relación entre el tema y la opinión suministrada; en cuanto que la opinión expresada puede diferenciarse, dependiendo del tema objeto de la opinión. (Hagedorn, Ciaranita, & Atserias, 2007) ilustran esta relación presentando las oraciones hipotéticas “*Walmart reports that profits rose*” y “*Target reports that profits rose*” cuya polaridad positiva o negativa estará determinada por el tema del documento, *Walmart*.

En cuanto al Análisis de Sentimientos Predictivo (Kim, Pantel, Chklovski, & Pennacchiotti, 2006) trabajan este enfoque con una mejoría de la precisión a la hora de clasificar el documento. Desde la perspectiva lingüística, los trabajos anteriores comprenden los principales avances en el área del AS, donde se busca representar mediante patrones de análisis documentos susceptibles de ser clasificados como positivos, negativos o neutros.

#### **2.4.2. ANÁLISIS DE SENTIMIENTOS DESDE LA PERSPECTIVA COMPUTACIONAL**

Desde una perspectiva computacional, el AS se mueve en los límites de los programas que inicialmente fueron diseñados para la recuperación de información por clustering. En los modelos de tópico, el AS ha encontrado acercamientos importantes a la solución de los problemas propios del campo. A continuación, se presenta una recopilación de los enfoques más importantes desde los modelos de aprendizaje no supervisado y supervisado.

##### **2.4.2.1. ANÁLISIS DE SENTIMIENTOS CON MODELOS DE TÓPICOS**

En principio, los modelos de tópico fueron desarrollados con la finalidad de descubrir y anotar (marcar) cantidades considerables de información temática, a escalas que serían imposibles de marcar manualmente. Estos modelos realizan análisis léxicos (palabras) para identificar los temas tratados en los documentos, las relaciones y los giros temáticos presentes en los mismos. Una de las características relevantes de estos modelos es que no requieren ningún tipo de marcación o etiquetaje lingüístico ya que el análisis surge de la misma información procesada. En la actualidad, se pueden clasificar los modelos de tópico bajo dos enfoques importantes: los modelos no supervisados y los modelos supervisados.

En los siguientes numerales se mencionan algunos de los trabajos más relevantes en AS bajo estos dos enfoques, ya que el objeto de nuestro trabajo de investigación está más orientado a la descripción lingüística que a la computacional.

#### ❖ ANÁLISIS DE SENTIMIENTOS A PARTIR DEL APRENDIZAJE NO SUPERVISADO Y APRENDIZAJE SUPERVISADO

Bajo las bases del modelo *Latent Dirichlet Allocation* (LDA) (Blei et al., 2003), diseñado para la agrupación de grandes volúmenes de información por temas, algunos investigadores han realizado importantes avances en la solución del problema de AS. Es importante resaltar que los modelos a partir del aprendizaje no supervisado no fueron diseñados en principio para AS, pero representan un avance importante en la solución del problema al permitir la construcción de vectores de características, como es el caso del modelo *Auxiliary-Sentiment Latent Dirichlet Allocation* (AS-LDA) (Liang, Liu, Tan, & Bai, 2014), el *Latent Dirichlet Markov Allocation* (LDMA) (Bagheri, Saraee, & de Jong, 2013), el *Dependency-Sentiment LDA* trabajado por (Li, Huang, & Zhu, 2007) basado en LDA; y, por último, el modelo *Dependency-topic-Affets-Sentiment-LDA* (DTAS-LDA) (Yin, Han, Huang, & Kumar, 2014). Al igual que con los modelos arriba listados, se presentan algunos ejemplos de modelos de AS a partir de aprendizaje supervisado importantes en el campo de trabajo como el modelo *Supervised Latent Dirichlet Allocation* (SLDA) desarrollado por (Blei & McAuliffe, 2010) y el modelo *multi-class sLDA* de (Wang, Blei, & Fei-Fei, 2009).

A pesar de la publicación, en años recientes, nos referiremos solamente a los modelos anteriores teniendo en cuenta nuestros objetivos de investigación. En conclusión, desde el enfoque *bag-of-words* hasta los n-gramas, las investigaciones en el campo han aportado elementos metodológicos fundamentales en la construcción de los vectores de características que permitan dar solución a los problemas de AS. Elementos lingüísticos como la negación, las conjunciones, las partes del habla y los recursos léxicos han enriquecido el ámbito metodológico de trabajo.

## **CAPÍTULO 3: DEL VERBO EN ESPAÑOL PARA PROPÓSITOS DE ANÁLISIS DE SENTIMIENTOS**

“Sin el verbo, el idioma sería un conjunto de interjecciones, gritos de aviso, sustantivos inconexos...” Alex Grijelmo (2006)

El presente capítulo pretende describir el verbo como unidad de estudio para propósitos de Análisis de Sentimientos.

Luego de presentar una mirada global sobre el AS, es posible observar, desde una perspectiva lingüística, el poco tratamiento que se le ha dado al verbo como unidad de análisis, en contraste con el adjetivo que ha sido ampliamente estudiado. Nos hemos planteado la tarea de describir el verbo con miras a contribuir en la construcción de la solución de los problemas inherentes al AS tanto en los aspectos metodológicos como teóricos.

En primer lugar, se presenta el verbo tanto en su descripción como en algunos aspectos de carácter general. En segundo lugar, se plantea cómo, desde los niveles morfológico, sintáctico y semántico de la lengua se concibe el verbo, para luego presentar, en un tercer momento, una tipología del mismo desde una perspectiva general y otra aplicada a los lenguajes de especialidad, ambas tipologías complementarias para los objetivos de nuestra investigación. En el momento final, se abordan los complementos argumentales como eje clave en la construcción de una caracterización del verbo y sus complementos para potenciar el análisis de sentimientos.

### 3.1. EL VERBO EN ESPAÑOL PARA PROPÓSITOS DE ANÁLISIS DE SENTIMIENTOS

Del latín *verbum* “palabra”, el verbo se define como la clase de palabras cuyos elementos pueden tener variación de persona, número, tiempo, modo y aspecto; y que constituyen el paradigma del verbo (Real Academia Española, 2017).

A través de la historia, el verbo ha sido objeto de diversos acercamientos y perspectivas teóricas, algunas complementarias y otras contradictorias. Como lo señala (Zamorano, 2002), para algunos teóricos como Pottier o Benveniste, el verbo es considerado el elemento fundamental en la construcción de las estructuras oracionales; mientras que otros teóricos como Hjelmslev niegan el papel nuclear del verbo en las proposiciones (p. 205).

Con relación a la definición del verbo, González Calvo (1978) presenta un recorrido en lo que concierne a las distintas formas en que se ha visto el verbo a través de la historia.

Desde la antigüedad, el verbo fue abordado por Platón desde el concepto *rhema* (predicado) en el sentido de predicado lógico. Aristóteles, por su parte, lo asume como término de predicación y determinante de tiempo desde del concepto de *onoma* (nombre).

Posteriormente, los estoicos retoman la idea platónica del verbo, dejando de lado su concepción temporal introducida en Aristóteles, la cual, según ellos, haría parte de la función semántica de los tiempos verbales del griego (Robins, 1997).

Con Dionisio de Tracia, nace la definición clásica del verbo al combinar criterios tanto morfológicos como sintácticos: “parte de la oración sin inflexiones de caso, que admite las de tiempo, persona y número y que expresa una actividad o un estado afectivo” (Carvajal, 1992, p. 3)

En la tradición latina, algunos gramáticos como Varrón, Donato, Prisciano y Consentio retoman el concepto de *verbum* desde los fundamentos alejandrinos.

Hacia el siglo XIII, los gramáticos escolásticos definieron el verbo conforme a la descripción latina de Prisciano, a pesar de ciertas conceptualizaciones en cuanto al modo de significación de

las palabras: “*Verbum est pars orationis cum temporibus et modis, sine casu. Agendi vel patiendi significativum*” (García, 1996).

Las primeras gramáticas del español, italiano y francés, publicadas en los siglos XV y XVI, retoman los elementos del verbo de Donato y Prisciano sin aportar nada nuevo a la construcción del concepto de verbo.

Antonio de Nebrija (1492:2010) en su gramática, afirma que el verbo en español es capaz de constituir por sí solo una oración, introduciendo el carácter predicativo del mismo:

Verbo es una de las diez partes de la oración, el cual se declina por modos et tiempos, sin casos. E llámase verbo, que en castellano quiere dezir palabra, no por que las otras partes de la oración no sean palabras, mas por que las otras sin ésta no hazen sentencia alguna, ésta, por ezcelencia, llamóse palabra. Los accidentes del verbo son ocho: especie, figura, género, modo, tiempo, número, persona, conjugación (capítulo X).

Pasando por la *Util y Breve Institución para Aprender los Principios y Fundamentos de la Lengua Española* y las definiciones en la gramática de Villalón (1558) o Gonzalo Correas, son retomados los elementos clásicos en cuanto a la combinación de elementos morfológicos, semánticos y sintácticos, en la definición del verbo.

Continuando con González Calvo (1978), J.C. Scaligero y el Brocense (como se le conocía a Francisco Sánchez de las Brozas), rechazan todo elemento distinto a los criterios morfológicos en la conceptualización del verbo, excluyendo en sus definiciones todo elemento con relación a la significación de las palabras (1978, p. 69).

El movimiento cultural e intelectual “La Ilustración” en el siglo XVIII, da entrada a la filosofía del lenguaje en la definición del verbo como categoría del habla. Las gramáticas filosóficas influenciadas por el movimiento racionalista, en especial las publicadas por las escuelas francesas de Port Royal, establecen la diferencia con los conceptos clásicos, al imponer la razón sobre la autoridad. Se retoman las ideas de Descartes como base en la concepción del verbo, más allá de lo trabajado por Aristóteles. Esta reformulación del concepto de verbo parte de la aceptación de las categorías de palabras clásicas, pero estas son agrupadas en palabras que significan los “objetos” del pensamiento y las que significan la “manera” de los pensamientos (González Calvo, 1978, p.

70). De esta forma, el verbo clasificado en el segundo grupo es definido como palabra, cuyo principal uso es significar afirmación; es decir, no sólo se conciben las cosas, sino que también se juzga sobre ellas y se afirma (Arnauld, 1846).

La Gramática Académica del siglo XVIII también fue influenciada por este movimiento. De allí que el verbo se defina como “parte principal de la oración que sirve para significar la esencia, la existencia, la acción, pasión y afirmación de todas las cosas animadas e inanimadas, y el ejercicio de qualquiera facultad que tienen estos casos o se les atribuye” (Real Academia Española, 1771).

En el siglo XIX, con la definición de Andrés Bello, (capítulo II, numeral 40, de su gramática), se tiene una perspectiva funcional del verbo, al definirlo como la “palabra que denota el atributo de la proposición, indicando juntamente el número y persona del sujeto y el tiempo del mismo atributo” (Bello, 1887:2014).

En la primera mitad del siglo XX, la Academia incluye criterios de tipo morfológico y lógico en su definición en la edición de 1931: “parte de la oración que designa estado, acción o pasión, casi siempre con expresión de tiempo y persona” (RAE, 1931).

En Gili Gaya (1980) el verbo se considera como un componente semántico de la oración que expresa la actividad o estado de un sustantivo.

En este siglo, se retoman los criterios semántico y funcional del verbo, planteando nuevas perspectivas unidas a un uso, en algunos casos, exclusivo de criterios morfológicos como en la gramática de la Academia del año 1973:

el verbo, por sus caracteres formales, es aquella parte de la oración que tiene morfemas flexivos de número, como el nombre y el pronombre, morfemas flexivos de persona, como el pronombre personal, y además, a diferencia del nombre y del pronombre, morfemas flexivos de tiempo y de modo. Suele aplicarse la denominación de desinencias a los morfemas de número y persona, el de características a los de modo y tiempo. Suprimidas de una forma verbal desinencias y características, lo que queda es la raíz o radical verbal. La agrupación de la raíz con la característica recibe el nombre de tema modal o temporal

el verbo, por sus caracteres formales, es aquella parte de la oración que tiene morfemas flexivos de tiempo y de modo (p.249).

Posteriormente, se retoma una visión más conservadora en los trabajos de Benveniste (1950, citado en González Calvo, 1978) revalorizando el concepto de verbo en la construcción de enunciados asertivos.

En general, los criterios semánticos, funcionales y morfológicos no son suficientes para definir el concepto de verbo, lo que lo aleja de ser considerado dentro de los “universales lingüísticos” (González Calvo, 1978, p. 80).

Para los objetivos de nuestro trabajo de investigación, el concepto de verbo está ligado al papel de este como núcleo de la oración. Álex Grijelmo (2006) en su gramática descomplicada afirma que “el verbo constituye la fuerza de la oración, establece las posibles relaciones entre los demás componentes de ella, permitiendo el vínculo entre los mismos” (p. 340). El verbo se sostiene por sí solo, es autosuficiente. En lenguas como el español, el verbo no depende de otras categorías gramaticales para expresar ideas completas: ven, mira, jugamos...

En síntesis, siguiendo la Gramática Descriptiva de la Real Academia, entendemos el verbo como:

Si por flexión se entiende las variaciones de forma de una palabra para manifestar distintos significados gramaticales o funcionales, el verbo es la palabra flexiva por excelencia: por el número de significados, de tiempo, aspecto y modo (TAM), y de número y persona (NP); y por las variaciones que de tales significados pueden expresar las distintas formas de un verbo. El verbo es una clase de palabras que significan un evento, una acción, proceso o estado. (Alcoba, 1999, p. 4917).

En la siguiente sección se puntualizará la forma como se aborda el verbo desde distintos niveles de la lengua, desde su construcción (morfología), interacción (sintaxis) y significación (semántica).

### **3.1.1. EL VERBO EN LOS NIVELES DE LA LENGUA**

Entendiendo la lengua como sistema (De Saussure, Bailly, & Séchehayé, 1919:1967), esta cuenta con una estructura ascendente que analiza los distintos elementos que la componen. Desde los fonemas hasta las estrategias de cohesión y coherencia, se han realizado un sinnúmero de estudios con relación a sus distintos elementos.

En los siguientes numerales, se abordará el verbo desde los niveles de la morfología, la sintaxis y la semántica para propósitos de AS.

#### **3.1.1.1. MORFOLOGÍA**

La morfología estudia la estructura de las palabras, las variantes que estas presentan y el papel gramatical que desempeña cada segmento con relación a los demás elementos que la componen (Real Academia Española, 2010). Este es el nivel de análisis cuyo objeto de estudio es el monema como unidad mínima de significado y de su subdivisión en lexemas y morfemas.

La morfología suele dividirse en dos grandes ramas: flexiva y lexical. Según la perspectiva de análisis puede ser sincrónica y diacrónica. En el caso de la flexiva, se estudian las variaciones de las palabras que implican cambios de contenido a nivel gramatical y, por consiguiente, en las relaciones de tipo sintáctico (Real Academia Española, 2010). La flexión del verbo (conjunto de cambios) aparece solo cuando la concordancia lo demanda y es exclusiva para determinar el modo, tiempo y aspecto.

Desde la morfología léxica, el verbo se concibe como una estructura bimembre: compuesta por una raíz como parte indivisible y las desinencias como los elementos que se añaden a la raíz, para precisar las variaciones de persona, número, tiempo, modo y aspecto (accidentes verbales) (Carvajal, 1992).

#### **3.1.1.1. PERSONA Y NÚMERO**

Con los morfemas de persona y número se establece la concordancia entre los núcleos del sintagma nominal (sujeto) y el del sintagma verbal (predicado). Con la persona gramatical se determina el proceso implicado en el verbo, es decir, se atribuye al hablante (primera persona), al



oyente (segunda persona) o al mundo aludido (tercera persona); mientras el número responde a la individualidad o colectividad de estos. En 1 y 2 se ilustra esta relación.

1. Los **pobladores** de San Basilio no **cuentan** con agua totalmente potable, el servicio de energía eléctrica es deficiente y costoso y carecen de alcantarillado.
2. Los **conflictos sociales**, asociados al marginamiento y la pobreza, **agudizan** las tensiones y **conforman** el escenario ideal para la actuación de los grupos violentos.

#### 3.1.1.1.1. TIEMPO

La noción de tiempo es propia del verbo recae y determina en qué momento se lleva a cabo el proceso de la oración: anterior (pasado), simultáneo (presente) o posterior (futuro). Esta noción (Benveniste, 1966:1977, p. 183) se refiere a que el tiempo cronológico coincide con el momento de la enunciación, salvo algunos casos particulares en los cuales el momento de la enunciación y la recepción son distintos.

El tiempo verbal consta de seis formas que varían conforme a la persona y número: primera, segunda y tercera persona tanto del singular como del plural. Los ejemplos 3, 4 y 5 corresponden al tiempo verbal.

3. La pobreza se **mantuvo** estable en Costa Rica (17,8 %), El Salvador (45,3 %), Uruguay (5,9 %) y República Dominicana (41,2 %), mientras que en México **aumentó** del 36,3 % a 37,1 %.
4. El departamento **ocupa** el octavo lugar en el país de familias que culminaron su proceso de acompañamiento.
5. Las cooperativas y las organizaciones de productores **tendrán** cada vez mayor importancia en el esfuerzo para eliminar el hambre y reducir la pobreza en el mundo [...]

#### 3.1.1.1.2. MODO

En español existen tres modos (indicativo, subjuntivo e imperativo) que expresan la relación del hablante frente al verbo. En otras palabras, manifiesta la actitud del hablante con lo que enuncia, cómo este percibe el mundo (Grijelmo, 2006, p. 218).

El indicativo, sin importar la temporalidad en que el enunciado es emitido, tiene que ver con la expresión de la realidad, o al menos, con lo que el hablante concibe como real. Es pues, la forma objetiva como se ve la realidad. En 6 y 7 se presentan ejemplos de oraciones en modo indicativo.

6. Uribe no sólo le devolvió la seguridad y la libertad al pueblo colombiano.
7. Es así como en las zonas periféricas de las capitales conviven la miseria, el hambre, el desempleo, la apatía, la injusticia, el resentimiento, el dolor y la carencia de acciones del Estado [...]

El subjuntivo, en oposición al indicativo, tiene que ver con la expresión de lo irreal, lo hipotético, lo dudoso; de aquello que existe solo en la mente del hablante. Los ejemplos 8 y 9 ilustran este modo.

8. Porque si fuera verdad que les interesara cambiar el planeta, ya se hubieran puesto a la tarea de erradicar la pobreza con su dinero, el que les sobra, y hace que existan cada vez más pobres.
9. Ojalá que la paz no necesitara de nadie.

El imperativo es utilizado para expresar ordenes, advertencias, amenazas o ruegos. La construcción de este modo tiene forma propia en la segunda persona tanto del singular como del plural, para las demás personas se recurre al subjuntivo.

Otros elementos importantes con relación al imperativo tienen que ver con la ausencia de la primera persona del singular, puesto que, gramaticalmente, no es lógico darse órdenes a uno mismo. Este modo se limita al presente y al futuro y se utiliza solo en oraciones independientes, es decir, el imperativo no se subordina ni subordina sintácticamente (Grijelmo, 2006).

### **3.1.1.1.3. ASPECTO**

El aspecto tiene relación directa con el tiempo. A diferencia de las lenguas eslavas o semíticas, en las lenguas de origen románico los morfemas que constituyen el aspecto verbal, son los mismos del tiempo. De allí que estos se confundan como una sola desinencia verbal (Castro & Rodríguez, 1974, p. 45).

Al respecto, varios autores han aportado a la construcción de una definición acerca del aspecto verbal. Este puede entenderse con relación al desarrollo del proceso verbal y se expresa por medios

tanto morfológicos como sintácticos, gracias a su carácter exclusivamente gramatical (Castro & Rodríguez, 1974, p. 45).

Sintácticamente, el aspecto en español es también expresado por medio de las perífrasis verbales, entendidas como toda expresión verbal no simple, compuesta de un elemento gramaticalizado (que aparte de tomar distintos morfemas, funciona en sí mismo como uno) y otro elemento no conjugado. (Castro & Rodríguez, 1974, p. 42). Por último, el aspecto determina si el proceso está o no finalizado. En 10 y 11, se presentan casos de aspecto verbal.

10. Estoy empeñado en buscar la paz para este país, y así como no he flaqueado y no me he debilitado en el propósito de luchar contra la pobreza para mejorar nuestra educación y empleo [...]

11. Conocí la problemática ambiental del departamento en ese ejercicio y al final la representante me dijo que le gustaba como trabajaba, que tenía mucho compromiso y que ella no iba a aspirar al Congreso.

### **3.1.1.2. SINTAXIS**

Es principio es necesario definir la oración, teniendo en cuenta que es el objeto de estudio de la sintaxis. El concepto de oración en Gili Gaya (1980) está construido a partir de tres puntos de vista (psicológico, lógico y gramatical) que se complementan para llegar a definir la oración como: “todos los elementos, palabras, frases u oraciones enteras, que se relacionen de modo inmediato o mediato con un verbo en forma personal” (p.23). De este modo, se puede establecer a nivel sintáctico las principales características del verbo y las relaciones de este con el resto de los elementos oracionales.

Con relación a esta triple clasificación de la oración, (Oronich, 2009) presenta una síntesis de cada aspecto en términos de la oración como unidad intencional (punto de vista psicológico), unidad de significación (punto de vista lógico) y unidad formal (punto de vista gramatical). Retomando los postulados de Husserl (1932) y Vossler (1940)<sup>1</sup>, autores en los cuales Gili Gaya

---

<sup>1</sup> Véase Husserl, *Investigaciones lógicas* (Publ. de la “Revista de Occidente”, Madrid, 4 Vols.); y Vossler, *Filosofía del lenguaje*, trad. de A. Alonso y R. Lida, Madrid, 1940.

(1980) basa su propuesta de la oración: la psicológica como algo notificado, la lógica como algo significado y, por último, la gramatical como algo nombrado.

La oración psicológica, corresponde a lo expresado; es decir, una idea o abstracción mental susceptible de ser expresada por medio del uso del lenguaje. Este tipo de oración es comprensible, aunque sea incompleta lógica o gramaticalmente.

La oración lógica se define como la expresión verbal de un juicio, el cual se entiende como la relación lógica entre un sujeto y un predicado. En toda oración se expresa una idea sobre alguien o algo. En consecuencia, tradicionalmente se ha definido el sujeto como la persona o cosa de la que se dice algo (tema) o bien la persona o cosa que ejecuta la acción del verbo (agente); y el predicado, como aquello que se dice del sujeto (rema) o la acción ejecutada por este (acción) y de allí las relaciones de tema-remata, agente-acción propias de la interacción sujeto y predicado.

En cuanto a la oración gramatical, esta se concibe como una unidad sintáctica, cuyo núcleo es un verbo conjugado. Al respecto, Gili Gaya afirma:

Observemos, finalmente, que los conceptos psicológico, lógico y gramatical de la oración, tal como han sido establecidos en este capítulo, se suman sin oponerse entre sí y se completan mutuamente. La unidad intencional centra la oración y la limita. La oración se organiza internamente con arreglo a valores psíquicos, entre ellos, y preferentemente, las leyes lógicas del juicio; y por último la expresión gramatical se articula en torno al verbo (p.25).

En esta línea, cuando en la oración existe un solo sujeto y predicado, se le atribuye el nombre de simple; mientras la combinación de dos o más sujetos y predicados se concibe como una oración compuesta. A partir de una doble forma de expresión, las oraciones tanto simples como compuestas, pueden ser clasificadas de diversas maneras. En primer lugar, la actitud del que habla; y segundo lugar, la naturaleza del sujeto y del predicado (Gili Gaya, 1980).

La figura 1 presenta la clasificación de Gili Gaya (1980) bajo estos dos criterios.

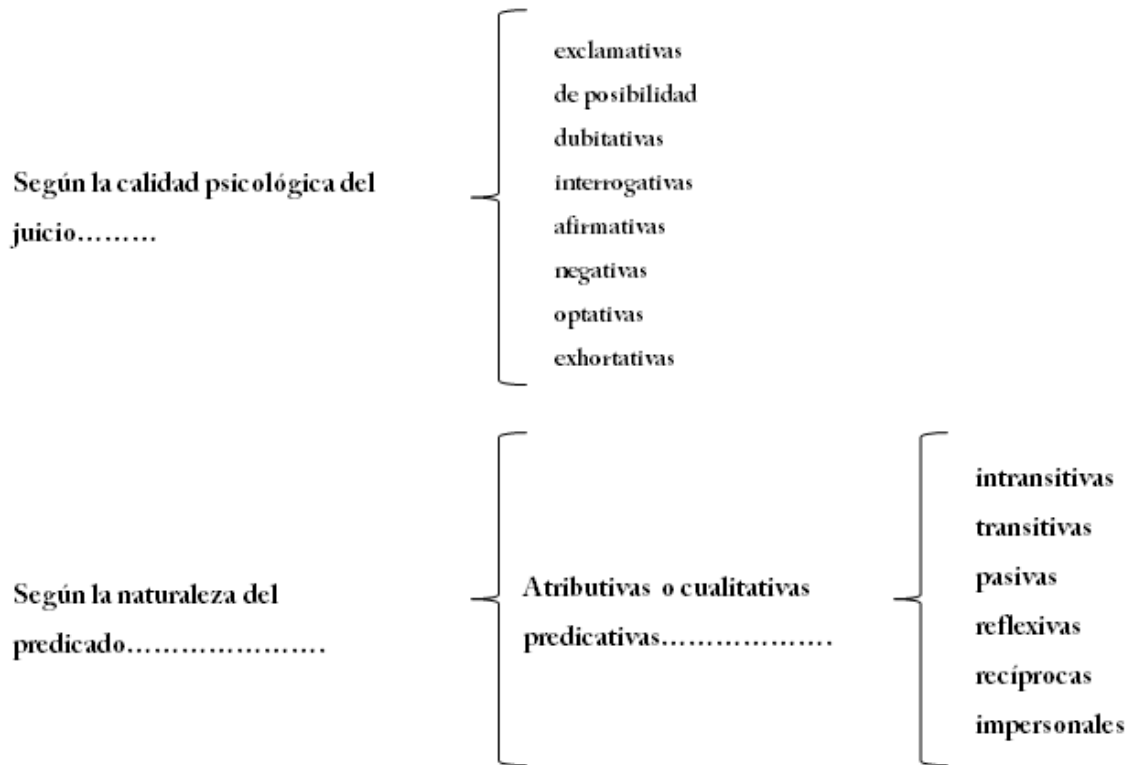


Figura 1: Tipos de oración en Gili Gaya

### 3.1.1.2.1. TEORÍA DE LAS VALENCIAS VERBALES

Fundada en los trabajos de Lucien Tesnière (1959), la teoría de las valencias verbales establece la dicotomía entre actantes y circunstantes. García-Miguel (2012) retoma la analogía que establece Tesnière con los elementos químicos:

*On peut comparer le verbe a une sorte de atome crochu susceptible de exercer son attraction sur un nombre plus ou moins élevé d'actants, selon qu'il comporte un nombre plus ou moins élevé de crochets pour les maintenir dans sa dépendance. Le nombre de crochets que présente un verbe et par conséquent le nombre d'actants qu'il est susceptible de régir, constitue ce que nous appellerons la valence du verbe (p.238).*

Esta teoría establece la cantidad de actantes y las relaciones sintácticas y semánticas implicadas; de allí que se hable de valencia cuantitativa o lógica, valencia sintáctica y valencia semántica siguiendo la propuesta de Helbig (1992) en *Probleme der Valenz- und Kasustheorie*.

Los actantes son entendidos como aquellos complementos exigidos por el lexema verbal; mientras, a los circunstantes pertenecen las adiciones innecesarias a la predicación (Crego, 2003, p. 69). Según Tesnière, el primer actante corresponde al sujeto, el segundo, es el complemento directo y el tercero sería el complemento indirecto. De acuerdo a su número, se determina la clasificación de los verbos según su valencia como nulos, monovalentes, bivalente, trivalentes y en algunas construcciones sintácticas podrían encontrarse verbos cuatervalentes.

Los verbos nulos o con valencia cero son aquellos que carecen de actantes; como es el caso de los verbos de clima, en los cuales no hay presencia de ningún otro elemento ni siquiera el sujeto de la oración: nevar, llover, escampar, helar, entre otros. Los verbos monovalentes son aquellos que solo reciben el primer actante (sujeto) como el caso de dormir, nacer, aparecer, entre otros.

Por otro lado, los verbos bivalentes, además de recibir el primer actante, también aceptan el segundo (complemento directo): verbos transitivos del tipo odiar, poseer, percibir, recordar, dibujar, reparar, entre otros; y verbos intransitivos del tipo molestar, suceder, incumbir, servir entre otros. Los verbos trivalentes, además del sujeto, aceptan los complementos tanto directo como indirecto. Algunos verbos de este tipo son estar, hacer, dar; además de verbos de asignación como reconocer o verbos de transferencia como prometer, decir, ordenar.

Algunos verbos son considerados cuatervalentes como traducir o comprar. Este último, en construcciones del tipo *yo compro un almuerzo a María por 10 mil pesos*. Las nociones en torno a la teoría de valencias han sido objeto de aplicación en distintos ámbitos de la lingüística general, en especial, en la lingüística de corpus, la lexicografía y el Procesamiento de Lenguaje Natural (García-Miguel, 2012, p. 30).

Finalmente, Schulte im Walde afirma que: “en procesamiento del lenguaje la mitad de los fallos en *parsing*, están causados por información valencial inadecuada” (2008, p. 954). Entendido *parsing* como “*the process of analysing the syntactic structure of a text or part of a text (such as a sentence)*” (McEnery & Wilson, 2001, p. 248)

### **3.1.1.3. SEMÁNTICA**

La semántica se entiende como la disciplina que estudia el significado de las unidades lingüísticas y de sus combinaciones (Real Academia Española, 2010). Este nivel estudia “el modo

como se proyectan los objetos y situaciones del mundo en el código de la lengua” (Helena, 2016, p. 2). En los siguientes numerales, se presenta una tipología verbal, teniendo en cuenta tanto elementos semánticos como sintácticos.

### **3.1.1.3.1. TIPOLOGÍA VERBAL**

Basados en los fundamentos tanto sintácticos como semánticos de los verbos, se presentan tres propuestas de clasificación: la primera, tiene en cuenta la expresión e interpretación de sus objetos directos (Levin, 1993); la segunda, es el proyecto *WordNet* que retoma la propuesta de (Levin, 1993); y por último, el postulado (Lorente, 2002) con relación a la función de los verbos en el discurso especializado. Por último, se presentan otras propuestas de clasificación verbal en los trabajos de Fillmore (1967) y Vendler (1967).

En Levin (1993), se encuentra la descripción de las categorías en la cuales se clasifican los verbos, de acuerdo a criterios de orden semántico. En el proyecto *WordNet*, se retoma la clasificación verbal de Levin (1993) y se presenta una propuesta de clasificación desde la sintaxis y la semántica, donde se tiene en cuenta los criterios de asociación entre los verbos. Estas categorías de clasificación, son las utilizadas en nuestra investigación para definir el sentido del verbo, a partir de los datos del corpus de trabajo.

En cuanto a la tipología ofrecida por Lorente (2002), sus categorías permiten identificar el papel desempeñado por cada uno de los verbos del corpus, considerando que estos están presentes en un discurso especializado: pobreza en Latinoamérica, específicamente en Colombia. En los siguientes numerales, se presenta cada una de estas descripciones, las cuales fueron aplicadas a la muestra de datos de la investigación.

#### **3.1.1.3.1.1.LEVIN (1993)**

En *English Verb Classes and Alternations*, Levin (1993) presenta una clasificación de los verbos partiendo de la concepción de que el comportamiento de un verbo está determinado por su significado con respecto a la expresión e interpretación de sus objetos directos. Algunas consideraciones previas de la autora remiten a la concepción del lexicón (partiendo de la idea de Bloomfield, 1983) como un apéndice de la gramática y una lista de irregularidades. Esta idea de léxico se refuerza al tomar el verbo como unidad de análisis, cuando el hablante demuestra un

amplio conocimiento lexical, al realizar juicios sutiles sobre la ocurrencia de los verbos. Estos presentan un amplio rango de posibles combinaciones debido a la habilidad del hablante para determinar el comportamiento del verbo con base en su significado (Levin, 1993).

En la clasificación de Levin (1993), se retoman aspectos tanto sintácticos como semánticos unidos a conceptos de alternancia. Según la autora, es posible agrupar en categorías semánticas los verbos, considerando una serie de diátesis. En su propuesta, estos son organizados en tres grupos conforme a las alternancias: en el primer grupo, se organizan las alternancias transitivas; mientras que, en el segundo, se reúnen aquellas que no afectan la transitividad. Por último, el tercer grupo contiene los verbos que permiten relaciones oblicuas en su clasificación.

El proyecto *WordNet* retoma esta categorización ofreciendo un conjunto de etiquetas con respecto al sentido de los verbos recopilados en su base de datos.

### 3.1.1.3.1.2. CLASIFICACIÓN VERBAL EN WORDNET

*WordNet* es una amplia base de datos lexical creada por el Profesor George Miller de la Universidad de Princeton. Está compuesta por 147.249 entradas que incluyen sustantivos, verbos, adjetivos y adverbios, agrupados en conjuntos de sinónimos cognoscitivos denominados *synsets*, expresados de manera individual e interconectados entre sí por medio de relaciones concepto-semánticas y léxicas.

En la tabla 1, se presenta la estadística de los datos generales de la versión de *WordNet* utilizada en nuestro trabajo de investigación.

Tabla 1: *Número de palabras, synsets y sentidos WordNet 2.1*

<b>Categoría léxica</b>	<b>Cadenas únicas</b>	<b>Synsets</b>	<b>Total pares palabra-sentido</b>
Sustantivo	117097	81426	145104
Verbo	11488	13650	24890
Adjetivo	22141	18877	31302
Adverbio	4601	3644	5720
Totales	155327	117597	207016

Los verbos se presentan y organizan de forma jerárquica, siguiendo las relaciones de troponimia, es decir, cuando un verbo se activa dentro de otro de alguna forma, como el caso de comunicar, hablar, susurrar. La base de datos también clasifica el modo de los verbos dependiendo



de las áreas temáticas: volumen, en el caso anterior; velocidad en moverse, trotar, correr; e intensidad de emoción en gustar, amar, idealizar. Para finalizar, los verbos que describen eventos que se implican necesaria y unidireccionalmente son vinculados entre sí: comprar-pagar, triunfar-intentar, mostrar-ver, entre otros. Es posible que una palabra o palabras se encuentren en más de un *synset* y categoría gramatical.

La figura 2 ilustra las relaciones verbales establecidas en *Wordnet*<sup>2</sup>.

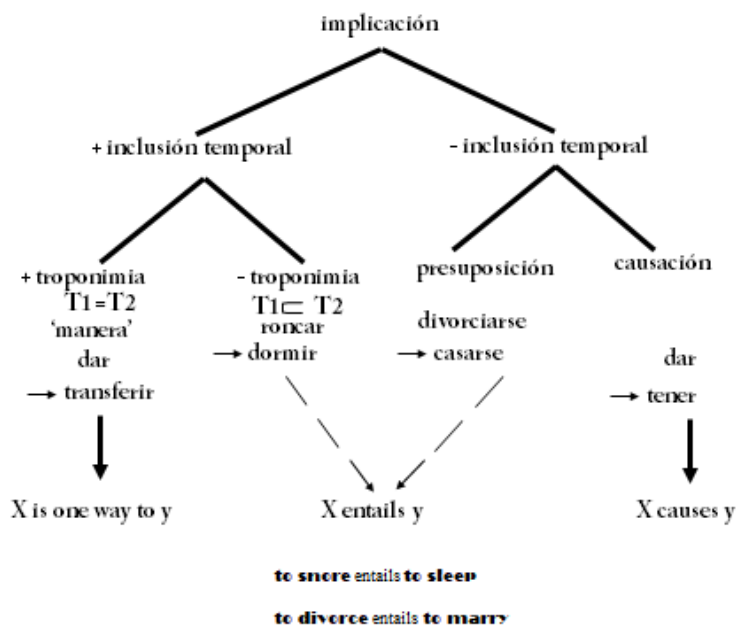


Figura 2: Relaciones verbales en *WordNet*

Un total de 15 categorías se utilizan en *WordNet* para agrupar los verbos como se muestra en la tabla 2.

Tabla 2: *Categorías de clasificación verbal en WordNet*

Name	Contents
verb.body	verbs of grooming, dressing and bodily care
verb.change	verbs of size, temperature change, intensifying, etc.
verb.cognition	verbs of thinking, judging, analyzing, doubting
verb.communication	verbs of telling, asking, ordering, singing

<sup>2</sup> Esta información fue recuperada de la documentación ofrecida en la página *Web* del proyecto. Véase <https://wordnet.princeton.edu/> para mayor información.

verb.competition	verbs of fighting, athletic activities
verb.consumption	verbs of eating and drinking
verb.contact	verbs of touching, hitting, tying, digging
verb.creation	verbs of sewing, baking, painting, performing
verb.emotion	verbs of feeling
verb.motion	verbs of walking, flying, swimming
verb.perception	verbs of seeing, hearing, feeling
verb.possession	verbs of buying, selling, owning
verb.social	verbs of political and social activities and events
verb.stative	verbs of being, having, spatial relations
verb.weather	verbs of raining, snowing, thawing, thundering

En los numerales 12 al 39 se listan algunos ejemplos recuperados del corpus de trabajo para ilustrar estas categorías, teniendo en cuenta que se hallaron 14 de los 15 sentidos (no hay datos con relación a los verbos de *weather*). Se conserva la denominación del sentido del verbo en inglés por razones metodológicas.

- Verbos de *body*

12. sufrir desnutrición crónica

13. no sufrir enfermedad

- Verbos de *change*

14. aumentar inversión social

15. mejorar condición de vida

- Verbos de *cognition*

16. reconocer proceso

17. saber camino

- Verbos de *communication*

18. explicar desastre

19. responder encuesta

- Verbos de *competition*

20. combatir pobreza extrema

21. superar dificultad fiscal

- Verbos de *consumption*

22. dar pescado

23. dar café

- Verbos de *contact*
- 24. no poner educación
- 25. poner investidura
  - Verbos de *creation*
- 26. construir vivienda social
- 27. generar trabajo digno
  - Verbos de *emotion*
- 28. preocupar pobreza
- 29. preocupar proceso del posconflicto
  - Verbos de *motion*
- 30. dejar tierra
- 31. llevar gas
  - Verbos de *perception*
- 32. presentar nivel de pobreza
- 33. sufrir violencia
  - Verbos de *possession*
- 34. comprar vivienda
- 35. ganar mínimo
  - Verbos de *social*
- 36. evitar abandono
- 37. trabajar inclusión social
  - Verbos de *stative*
- 38. existir empleo
- 39. vivir jornal diario

#### **3.1.1.3.1.3. LORENTE (2002)**

En el campo de la lexicografía especializada, al igual que en el AS, la mayoría de las investigaciones se han centrado en otras categorías gramaticales distintas al verbo, siendo este poco trabajado o en ocasiones excluido. Sin embargo, como afirma Lorente (2002), la realidad es otra. En el discurso especializado, el verbo es parte significativa, además de aportar información relevante que no puede omitirse.

Para Lorente, los verbos poseen características que deben tenerse presente dado que aportan elementos importantes en el análisis terminológico. Desde esta perspectiva, se pueden encontrar verbos:

- portadores de conocimiento específico
- de uso general que en contexto especializado presentan rasgos semánticos distintivos
- vinculados morfológicamente a otras unidades de categoría nominal o adjetival, con los que forma familias derivativas
- de uso general cuyo comportamiento sintáctico difiere cuando aparecen en textos de registro distinto.

Sobre estas características, Lorente (2002) concluye que en toda descripción del discurso especializado y en toda construcción de aplicaciones, debería existir un balance con relación a las unidades verbales. Además, la presencia de estas unidades permite realizar un análisis desde el punto de vista morfológico, sintáctico y semántico.

Desde varios ámbitos, existen motivaciones por las cuales el tratamiento verbal en el discurso especializado ha sido tan reducido. En su trabajo, Lorente las describe desde los aspectos: metodológicos de la técnica y la ciencia, aspectos gramaticales y aspectos pragmáticos o lexicográficos y, por último, en usos desde la lingüística aplicada.

Metodológicamente y de acuerdo con la Teoría General de la Terminología (Wüster, 2010), las categorías que han sido privilegiadas son de carácter nominal, considerando que el objetivo principal de esta, remite a la internacionalización de los términos técnicos y científicos. Gramaticalmente, el significado del verbo unido a los complementos no es suficiente para dar sentido completo de la predicación, debido a su carácter abierto, lo que ha facilitado que se recurra a categorías léxicas nominales.

Pragmáticamente, el carácter de especialidad del verbo es atribuido al contexto con el cual este confluye; esto es, al sujeto y/o complementos regidos, generando así una predilección por las unidades nominales donde esta característica se cumple en mayor proporción. Lexicográficamente, la metodología prioriza las definiciones por comprensión; es decir, basadas en un descriptor que represente la clase semántica de las palabras contenidas en la palabra definida y, a su vez, los

rasgos que la distinguen de las otras palabras de su clase; característica de difícil aplicación a las unidades verbales.

Desde el diseño y los objetivos de la Lingüística Aplicada, la finalidad de estos ha determinado el uso de las unidades léxicas cerradas en la fijación y unidades léxicas integrables fácilmente en jerarquías para las ordenaciones. En la actualidad, las aplicaciones para recuperación de información en la Internet o la gestión del conocimiento, demandan la necesidad de incluir verbos como unidades de tratamiento.

Lorente (2002) presenta una tipología de clasificación de los verbos en contextos especializados, haciendo énfasis en que es una propuesta que “solo persigue ofrecer algunas orientaciones para el diseño de aplicaciones más adecuadas” y que es una clasificación “dependiente de cada ámbito de especialidad, ya que un determinado lema verbal puede considerarse distintamente cuando se usa en un contexto temático o en otro” (par.28).

#### **3.1.1.3.1.3.1. VERBOS DISCURSIVOS**

Esta clase de verbos se relacionan con las estrategias propiamente textuales empleadas por el emisor, con las cuales se busca activar la competencia pragmática, pero no se activa ningún tipo de conocimiento especializado. Sin embargo, hay casos híbridos donde un verbo discursivo podría activar información relativa a la metodología de trabajo (hipotetizar, analizar, clasificar, deducir, entre otros).

Lorente (2002) identifica estos verbos como referentes a los actos de habla (decir, comunicar, transmitir, opinar); a las funciones textuales (describir, narrar, dar instrucciones, argumentar); a la estructura del discurso (organizar, estructurar, ordenar, continuar, concluir) o a la finalidad discursiva (presentar, convencer, instruir, demostrar); al igual que los verbos psicológicos, poco frecuentes en el discurso científico-técnico (pensar, opinar, temer, dudar). En suma, la autora considera importante hacer una indicación en las herramientas de tratamiento automático del lenguaje, donde se incluya la función meramente discursiva de este tipo de verbos.

#### **3.1.1.3.1.3.2. VERBOS CONECTORES**

Una segunda clasificación en Lorente, son los verbos conectores. Estos hacen parte de la expresión del conocimiento especializado, pero no lo proporcionan, puesto que su función, como

lo indica su denominación, es la de expresar relaciones de equivalencia, igualdad, similitud, dependencia o atribuir cualidades o valores. Pertenecen a esta categoría verbos como ser, parecer, equivaler, corresponder.

### **3.1.1.3.1.3.3. VERBOS FRASEOLÓGICOS**

Los verbos fraseológicos expresan acciones, procesos y estados en los contextos en los cuales confluyen con sus respectivos complementos directos; es decir, cuando forman parte de unidades sintagmáticas en las que adquieren valor especializado y, por lo tanto, transmiten este tipo de conocimiento, solo cuando aparecen como unidades terminológicas. Los verbos de soporte también son susceptibles de integrar esta categoría, toda vez que forman parte de combinaciones léxicas fijas.

### **3.1.1.3.1.3.4. VERBOS-TÉRMINO**

Por último, los verbos-término son aquellos ligados exclusivamente a un conocimiento especializado, siendo frecuentes las unidades formadas por derivación y las variantes sintácticas del verbo *hacer* más un derivado nominal. Por ejemplos: Capialzar (arquitectura), eutrofizar (ecología), acetificar (química). En esta categoría, el nivel de especialidad puede ser atribuido tanto al verbo como al sustantivo del cual se deriva, es el caso de demandar y demanda.

### **3.1.1.3.1.4. OTRAS PROPUESTAS DE CLASIFICACIÓN**

Al igual que las tres propuestas previamente descritas, es importante mencionar otros acercamientos a una descripción del verbo desde los niveles sintáctico y semántico que ofrecen más elementos en la descripción de las unidades verbales. A continuación, se presenta una breve mirada a dos propuestas de clasificación verbal en los trabajos de Vendler (1967) y Fillmore (1967).

En *Linguistics in Philosophy*, Vendler (1967) agrupa los predicados de las unidades verbales en 4 categorías, a saber: estado, actividad, realización y logro. En *The Case for Case*, Fillmore (1967) presenta una descripción de las unidades verbales, aludiendo que los aspectos formales no son suficientes para expresar las relaciones establecidas entre los elementos de la oración, por lo que requieren de la semántica, máxime cuando el verbo es el elemento central.

### 3.1.2. COMPLEMENTOS ARGUMENTALES

Siguiendo la relación entre el sujeto y el predicado, en las oraciones hay elementos que son indispensables y otros que cumplen una función exclusivamente semántica, al añadir unidades para brindar una mayor precisión contextual (Dapena, 2002, p. 7). Los complementos aportan información adicional al contenido de las unidades de las cuales dependen.

El hecho de que los complementos no afecten directamente la estructura de la oración, confirma su estatus como elementos opcionales de la misma, siendo esta característica exclusiva desde el punto de vista sintáctico, y teniendo en cuenta que, desde una mirada semántica, los complementos pueden llegar a ser obligatorios (Dapena, 2002, p. 7).

Otra característica de los complementos, es su papel como expansión de un núcleo sintagmático. No solo como expansión de este, sino también como unidad constitutiva del elemento que acompañan. En consecuencia, Dapena (2002, p. 10) se decanta por una denominación de estos como complementos de nombre, verbo, de adjetivo o de adverbio; capaces de conformar núcleos sintagmáticos. En el numeral 40, se observa esta característica en “política de vivienda” en contraste con el 41, donde se suprime el complemento de política, creando un sentido distinto al inicial, aunque no ambiguo.

40. "Por eso la política de vivienda le apuesta ahora a cambiar la vida de más colombianos pobres con un techo digno".

41. "Por eso la política le apuesta ahora a cambiar la vida de más colombianos pobres con un techo digno".

El concepto de complemento que se ha utilizado, corresponde a la propuesta de Dapena (2002), en oposición al punto de vista funcional de complemento, propuesto por Alarcos (1999). Por ende, siguiendo los objetivos planteados en la investigación, se trabajarán solo aquellos con relación al núcleo del predicado, que siempre corresponde a un verbo; en específico, el complemento directo, junto a una breve descripción de los demás tipos de complementos en la propuesta de Dapena (2002).

### 3.1.2.1. TIPOS DE COMPLEMENTOS ARGUMENTALES

Para Dapena (2002, p. 10), los denominados complementos verbales son heterogéneos y no se limitan a los tradicionales (directos, indirectos, circunstanciales y predicativos). Por otro lado, de acuerdo con el grado de incidencia sobre el núcleo predicativo, se establece una jerarquización de los complementos. La figura 3, muestra esta relación introducida en Dapena (2002, p. 16).

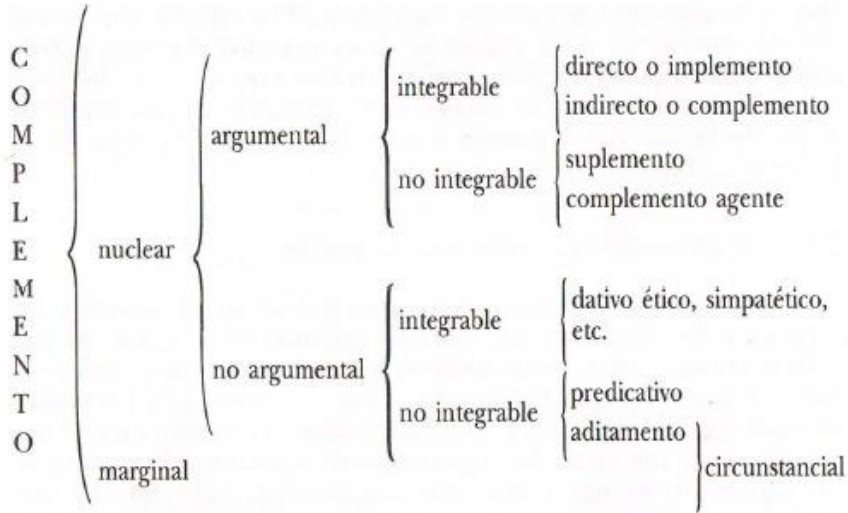


Figura 3: Tipos de objetos directos Dapena (2002)

En los siguientes numerales, hemos descrito cada tipo de complemento presentado en la figura anterior.

#### 3.1.2.1.1. COMPLEMENTOS NUCLEARES Y MARGINALES

Con relación al núcleo del sujeto y predicado, la importancia sintáctica de los complementos, en esta categoría, es relativa al grado de cohesión que se establece con el verbo; entre menor sea la cohesión, mayor es la movilidad al interior de la oración. Los complementos nucleares son aquellos que guardan una mayor cohesión con el verbo y al mismo tiempo, son susceptibles de ser enfatizados. Por otra parte, se define como complementos marginales, aquellos que presentan una menor cohesión.



### **3.1.2.1.2. COMPLEMENTOS ARGUMENTALES O REGIDOS Y NO ARGUMENTALES**

La segunda categoría corresponde a los complementos nucleares, dado que estos también afectan al verbo de distinta forma y están o no condicionados por elementos léxicos propios del verbo. Los complementos argumentales o regidos son aquellos exigidos por el verbo y su aparición depende de este, mientras la aparición de los no argumentales no está supeditada a las características léxicas del verbo y puede ser cualquier tipo de objeto directo.

En el caso de los verbos de acción, determinar si el complemento es argumental o no, se logra al sustituir el sintagma verbal por *hacerlo*. De ser positiva la sustitución, en términos de sentido se habla de un complemento no argumental y, en caso contrario, de uno argumental. Sin embargo, el criterio más acertado para lograr esta distinción consiste en un análisis semántico del verbo; información que pocos diccionarios ofrecen y que en este punto es de vital importancia. Tomando la segunda acepción en la definición del verbo comunicar, en el Diccionario de la Real Academia, se puede construir la siguiente estructura:

Comunicar: Descubrir, manifestar o hacer saber a [alguien] [algo].

Entre corchetes se encerraron los elementos que han de ser llenados por los complementos. Estos corresponden a los complementos argumentales y, la información adicional, es considerada como los complementos no argumentales. Adicionalmente, según Dapena (2002, p. 12), hay dos criterios para determinar que un complemento sea argumental. En primer lugar, si la viabilidad del enunciado depende del complemento, este será de tipo argumental y, en segundo lugar, frente a la presencia innecesaria del complemento para la viabilidad del enunciado. Para esta, hay dos posibilidades, primero que la construcción entre verbo y complemento esté mediada por los indefinidos *o*, segundo que ante la ausencia de este, se presente un cambio de sentido.

### **3.1.2.1.3. COMPLEMENTOS INTEGRABLES Y NO INTEGRABLES**

La tercera categoría se refiere a los complementos integrables o no. Para Dapena (2002, p. 12), se hace necesaria una tercera subclasificación con relación al uso de los pronombres clíticos o átonos. En este sentido, los complementos directos e indirectos pueden ser reemplazados por este tipo de pronombres. Los complementos de suplemento y los predicativos no son susceptibles

de ser reemplazados, de allí que los primeros se clasifiquen como integrables y los segundos como no integrables. En última instancia, la tipología descrita tiene fundamento tanto operacional como funcional. Todos los complementos descritos ocupan un lugar distinto con relación a las estructuras sintácticas (Dapena, 2002, p.12).

### **3.1.2.2. COMPLEMENTO DIRECTO O IMPLEMENTO**

Hablar de complemento directo u objeto directo implica, entre muchos elementos, problemas al momento de establecer una definición sobre qué es este complemento. El directo, es el primer complemento del verbo, en otras palabras, es el primer elemento que modifica o complementa al verbo (Dapena, 2002, p.18). No obstante, confluyen otros elementos tanto funcionales (aspecto semántico) como formales (aspecto sintáctico) que contribuyen a una definición más completa. La transitividad del verbo determina la existencia o no de este complemento, lo que a su vez implica solucionar el dilema de la transitividad.

En la *Sintaxis Histórica de la Lengua Española*, coordinada por Concepción Company Company, Brenda Laca (2006) aborda el complemento directo en la misma línea de Dapena, definiéndolo como “un objeto directo estructuralmente adyacente a una forma verbal activa que lo rige” (pp. 1–2). La construcción de una definición de complemento de objeto directo se fundamenta en elementos sintáctico-semánticos. En este orden de ideas, se aborda una descripción del complemento directo o implemento desde su caracterización tanto funcional como formal.

En primer lugar, el objeto directo, a nivel semántico, se define con relación a la transitividad del verbo, definida en torno al objeto directo. Esto plantea la necesidad de definir la transitividad como propiedad semántica por la cual un verbo expresa una acción o proceso, que va de un agente u origen hacia un paciente o termino (Dapena, 2002, p. 18). En el numeral 42, se ejemplifica esta relación.

#### **42. Parafiscales crean desempleo: Emisor**

Para Dapena (2002), el carácter transitivo no solo expresa la realidad de la acción, sino que sirve como puente entre los dos elementos, ya sea desde el agente (exocéntrica) o desde el paciente (endocéntrica). En consonancia, Laca (2006) expresa la notoria variación de esta relación en cuanto al tipo de verbo. En suma, esto aplicaría a los verbos de acción dejando por fuera los verbos

de estado. Además, de la confusión en los casos donde el agente recibe la acción del verbo y no el paciente. En los numerales 43 y 44, se ilustra esta confusión.

43. Los habitantes de Las Playas sufren con frecuencia de enfermedades como infecciones, malaria y tuberculosis que en ocasiones les lleva a la muerte.

44. Hoy, 50 años después, las consecuencias de este conflicto han superado con creces las desgracias ocasionadas al país por las mismas causas que dieron origen a los grupos guerrilleros.

Dapena (2002) presenta un segundo matiz en la definición del implemento. El objeto directo es abordado simplemente como un adyacente verbal, cuya función se limita a completar el sentido del verbo. Esto es, una suma de verbo transitivo más objeto directo, aunque de nuevo no es suficiente para definirlo. En cuanto a la función que le ha sido atribuida, es la misma de todos los complementos; además de la información semántica que no siempre es menor en los verbos transitivos, en contraste con los intransitivos (p.19).

En segundo lugar, debido a la evidente dificultad a la hora de definir el objeto directo desde aspectos funcionales, los lingüistas modernos recurren al aspecto formal. En este sentido, Dapena (2002) y Laca (2006) basan sus descripciones en rasgos morfológicos y distribucionales o en conmutaciones o transformaciones (pruebas operacionales). Para el español, ambos coinciden en que solo son aplicables las pruebas operacionales para definir el implemento.

Con relación al objeto directo, Dapena (2002) retoma en primer lugar, la condición integrable del implemento, es decir, este es susceptible de ser pronominalizado, al ser sustituido por las formas átonas del acusativo o clíticas (o, la, los, las) (p.20). En numeral 45, se ilustra esta característica al igual que la forma reemplazada.

45. (a) Programa Jóvenes en Acción entregará **becas**

(b) Programa Jóvenes en Acción las entregará

A su vez, la tematización del sintagma del objeto directo que, al anteponerse al verbo, exige un pronombre anafórico de acusativo (p.21). El numeral 46 da cuenta de ello al igual que ilustra la forma reemplazada.

46. (a) Programa Jóvenes en Acción entregará **becas**

(b) las becas **las** entregará el Programa Jóvenes en Acción

Otra condición es el cambio a pasiva del sujeto como se observa en el numeral 47.

47. (a) Programa Jóvenes en Acción entregará **becas**

(b) las becas **fueron entregadas por** el Programa Jóvenes en Acción

Este procedimiento también ofrece la posibilidad de concordar el verbo en participio con el objeto directo como en se muestra en el ejemplo 48.

48. (a) Programa Jóvenes en Acción entregará **becas**

(b) las becas **entregadas por** el Programa Jóvenes en Acción

Otra condición es la sustantivación del participio por medio del artículo neutro *lo*, corroborando que este se refiera al mismo objeto que el sintagma nominal enlazados por el verbo ser. En el ejemplo 49 se da cuenta de esto.

49. (a) Programa Jóvenes en Acción entregará **becas**

(b) **lo entregado** por el Programa Jóvenes en Acción **fueron** las becas

Por último, convertir la oración en pregunta por medio de ¿qué? ¿Qué es lo que? o ¿a quién? cuya respuesta será el objeto directo como se ve en el ejemplo 50.

50. ¿Qué entregará el Programa Jóvenes en Acción? R/ las becas

Sin embargo, según los criterios descritos aún persisten problemas en su utilización. Por un lado, los fenómenos de laísmo, leísmo y loísmo favorecen la confusión con la respuesta, al convertir la oración en pregunta en los casos donde el sintagma posee rasgo + humano. Esto es susceptible de responderse con el objeto directo o dativo; mientras que, por otro lado, la conversión a pasiva se ve afectada considerando que no todos los verbos transitivos pueden ser transformados a voz pasiva.

Siguiendo la línea de trabajo en Dapena (2002), los problemas de delimitación del implemento se plantean con relación a los complementos indirectos y algunos circunstanciales (medida o cantidad); y en ocasiones con los predicativos, teniendo como base la pronominalización como la estrategia más adecuada para la identificación del objeto directo.

El español, a diferencia de otras lenguas, carece de una marca distintiva en la identificación del objeto directo, lo que genera problemas en la delimitación del mismo. En efecto, se denomina objeto directo (siguiendo la tradición francesa) puesto que no recibe ningún tipo de modificación. Esta condición es parcialmente verdadera para el español, donde en ocasiones se introduce el objeto directo por medio de la preposición *a* y de allí, el uso de implemento como forma de denominación introducida por Alarcos desde el año 1968, dando origen a los dos tipos de implementos del español: el no preposicional o directo y el preposicional. Con relación a este aspecto, Laca (2006) plantea tres hipótesis, partiendo de tres autores, sobre el origen de lo que ella denomina el acusativo preposicional: en Müller (1971) como analogía con el sujeto, en Meyer-Lübke (1899) como analogía con el dativo y en Rohlfs (1971) con el estatus de tópico.

La función de esta preposición en la construcción del implemento directo es exclusivamente diacrítica, al permitir diferenciarlo del sujeto de la oración, olvidando la función relacionante propia de su naturaleza (Dapena, 2002, p. 23). Por lo tanto, esta puede ser interpretada como un sintagma nominal, con o sin la presencia de esta preposición. El ejemplo 51, presenta el uso de la preposición *a* en el implemento directo.

51. Germán Vargas Lleras [...] dice que le gustaría **ver a** Álvaro Uribe apoyando el proceso en vez de estar poniéndole palos en la rueda.

En adición, el implemento directo no siempre es obligatorio en la oración pese a su carácter argumental, lo que no implica que este tipo de oraciones sean consideradas intransitivas, siendo el objeto directo considerado como explícito o implícito (Dapena, 2002). Para concluir, existen otros tipos de objeto directo, tales como: el efectuado o causado; el afectado o modificado; y el complemento directo ni afectado ni causado, determinado por el efecto causado por la acción o proceso del verbo sobre el objeto del implemento directo.

### **3.1.2.3. OTROS TIPOS DE COMPLEMENTOS: INDIRECTO, SUPLEMENTO Y AGENTE**

En español, existen otros complementos argumentales del verbo, a saber: indirecto, de suplemento y agente. A continuación, se presenta una breve descripción de estos tres complementos del verbo.

#### **3.1.2.3.1. COMPLEMENTO INDIRECTO**

Los problemas al delimitar el complemento indirecto (antes denominado como el complemento verbal con preposición) son mayores que aquellos encontrados al delimitar el implemento. Al respecto, existen dos acercamientos contrarios a la idea inicial de “complemento verbal con preposición”.

En primer lugar, en la Gramática de la Real Academia de 1962, el indirecto es definido en términos del “vocablo que expresa la persona, animal o cosa en quien se cumple o termina la acción del verbo transitivo ejercida ya sobre el acusativo, y también la de verbos intransitivos” (p.239). En segundo lugar, el *Esbozo de una Nueva Gramática de la Lengua Española* del año 1973, define el complemento indirecto únicamente con relación a la transitividad.

En la definición de la *Nueva Gramática* (2010), se continúa con la línea presentada en el Esbozo de 1973. El complemento indirecto cumple sintácticamente con la función de dativo, el cual es trabajado formalmente con relación a su aparición, precedido de la preposición *a* o *para*. Sin lugar a dudas, esto es insuficiente en su concepción, considerando que el implemento también es introducido en ocasiones por la preposición *a*.

#### **3.1.2.3.2. SUPLEMENTO Y COMPLEMENTO AGENTE**

Con relación a los complementos circunstanciales, cuya denominación está definida en negación a complementos directos o indirectos (ni son directos ni son indirectos), Dapena (2002), utiliza la definición de Alarcos, quien define suplemento como el tipo de complemento circunstancial más cercano de ser categorizado como complemento directo preposicional; diferenciándolo de los aditamentos que serían los típicos modificadores adverbiales de modo, lugar, tiempo, cantidad, etc.

Con relación al suplemento, existen algunos criterios variables e inestables para su definición lo que dificulta su delimitación, al igual que con los complementos directos e indirectos. Alarcos (1968) propone una serie de criterios para definir el suplemento.

El primer criterio, corresponde a la imposibilidad del suplemento de ser conmutado con un adverbio, salvo mediante un pronombre personal tónico precedido de la misma preposición del enunciado. El segundo criterio de delimitación del suplemento, es la obligatoriedad e imposibilidad de ser eliminado de la oración. El tercer criterio, es la incompatibilidad de aparición frente al aditamento en la misma oración.

Al igual que con los criterios en los otros complementos, se presentan dificultades con relación a las formas de delimitación del suplemento. En última instancia, el complemento agente de carácter opcional construido por un sintagma preposicional con *por* o *de*, y la construcción pasiva del verbo transitivo, el cual también presenta problemas de delimitación y que aparece con mayor frecuencia en las construcciones con rasgo +potente (Dapena, 2002), es decir, capaz de obrar o desarrollar la acción contenida en el verbo.

# **PARTE II: METODOLOGÍA DE INVESTIGACIÓN**





# CAPÍTULO 4: METODOLOGÍA DE INVESTIGACIÓN

## 4.1. ENFOQUE METODOLÓGICO

Al definir el enfoque metodológico que hemos seguido en nuestro trabajo de investigación, tomamos como referencia, entre otras, a Hernández, Fernández y Baptista (2014). Nuestra investigación se enmarca dentro de un enfoque de investigación mixto, toda vez que combina elementos cuantitativos y cualitativos. Así mismo, la investigación mixta tiene como finalidad hacer uso de las fortalezas de ambos enfoques, tal como lo aclara Hernández *et al.*,

La meta de la investigación mixta no es reemplazar a la investigación cuantitativa ni a la investigación cualitativa, sino utilizar las fortalezas de ambos tipos de indagación, combinándolas y tratando de minimizar sus debilidades potenciales (p.532).

No obstante, nuestra investigación presenta una mayor tendencia hacia lo cuantitativo que hacia lo cualitativo, siguiendo lo afirmado por Jhonson, Onwuegbuzie y Turner (2007) en *Toward a definition of Mixed Methods Research* con relación al “*sentido amplio*” de la investigación mixta y los diferentes matices al combinar ambos enfoques. En este sentido, los datos cualitativos, durante la etapa del análisis de los datos, juegan un rol importante “*by interpreting, clarifying, describing, and validating quantitative results, as well as through grounding and modifying*”(Johnson *et al.*, 2007, p. 115).

Con relación a las patrones [verbo + objeto directo], los datos que hemos recolectado como unidades de análisis permiten, a partir de un “*conjunto de procesos sistemáticos, empíricos y críticos [...] lograr un mayor entendimiento del fenómeno bajo estudio*” (Hernández *et al.*, 2014, p. 534). En este orden de ideas, Johnson *et al.* (2007) plantea la dinámica de las investigaciones mixtas con tendencia cuantitativa, unida al aprovechamiento de las ventajas de la investigación cualitativa:

*Quantitative dominant mixed methods research is the type of mixed research in which one relies on a quantitative, postpositivist view of the research process,*

*while concurrently recognizing that the addition of qualitative data and approaches are likely to benefit most research projects (p.124).*

## **4.2. ALCANCE DE LA INVESTIGACIÓN**

El alcance de nuestra investigación es descriptivo; es decir, es una investigación que “busca especificar propiedades y características importantes de cualquier fenómeno que se analice”. Lo anterior nos permite “mostrar con precisión los ángulos o dimensiones de un fenómeno, suceso, comunidad, contexto o situación” (Hernández *et al.*, 2014, p. 92). En este sentido, este trabajo de investigación tiene un carácter predictivo, puesto que los resultados del análisis de datos puede servir para trabajos en diferentes áreas, en especial, la guía metodológica,

## **4.3. ENFOQUE BASADO EN CORPUS (*CORPUS- BASED*)**

Siguiendo con lo expresado anteriormente, para dar cumplimiento a nuestros objetivos y corroborar las hipótesis de trabajo, hemos definido un enfoque basado en corpus (*corpus-based*). Entendemos por corpus a la luz de la definición de Quiroz (2003),

un conjunto de textos (escritos u orales) legibles en un computador que representan el uso real del lenguaje en situaciones determinadas y se usan, principalmente para investigación y creación de materiales para diferentes áreas relacionadas con el lenguaje (p.133).

Un enfoque basado en corpus hace uso de la información en aras de explorar teorías o hipótesis existentes en la literatura con la finalidad de validarlas, refutarlas o refinarlas (McEnery & Hardie, 2012, p. 6). Este enfoque nos ha permitido aprovechar las ventajas de la lingüística computacional, aplicadas a los estudios léxicos, facilitando el análisis de combinaciones de palabras como lo expresa McEnery y Wilson (2001): “*to call up word combinations rather than individual words*” (p. 108). Además, los autores indican la existencia de herramientas para el procesamiento de los datos recopilados, como frecuencias, concordancias, correlaciones, listas de palabras clave, entre otras.

## 4.4. CONCEPCIÓN DEL CORPUS DE TRABAJO Y METODOLOGÍA DE TRABAJO

En el desarrollo de este tipo de investigación, la concepción y posterior elaboración del corpus de trabajo constituyen, sin lugar a dudas, el mayor de todos los desafíos. Para dar buen término a los objetivos trazados y abordar las hipótesis planteadas, se tomó como referencia el corpus de pobreza del proyecto PoLaMe del cual se seleccionó una muestra aleatoria estratificada y se siguió la siguiente metodología de trabajo.

### 4.4.1. DESCRIPCIÓN DEL CORPUS DE TRABAJO

En el marco de la presente investigación, se tomó como referencia el corpus de pobreza del proyecto PoLaMe: *Poverty, Language & Media*; un proyecto internacional multidisciplinar liderado por la Universidad de Bergen, Noruega y financiado por el *Norwegian Research Council* (Consejo Noruego de Investigación) en el período comprendido entre los años 2012 - 2017; el cual buscó analizar, desde la lingüística, el fenómeno de la pobreza en 4 países latinoamericanos (Argentina, Brasil, Colombia y México) teniendo como base artículos de prensa escrita publicados en los periódicos más representativos de cada país, entre estos: Clarín, La Nación, Página 12, La Voz del Interior, La Gaceta; de Argentina; Folha de São Paulo, O’Globo, Gazeta; de Brasil; El Tiempo, El Espectador, El Colombiano; de Colombia; y Excélsior, La Jornada, La Reforma; de México.

El corpus cuenta con más de 37 millones de palabras en español y portugués (alrededor de 30 mil artículos) y fue compilado automáticamente utilizando tecnologías lingüísticas modernas según la descripción de este ofrecida en la *Web* oficial del proyecto<sup>3</sup>.

La extracción de los artículos se realizó de forma automática, semiautomática o manual dependiendo de la viabilidad de los sitios web de cada periódico; es decir, si estos utilizan tecnología compatible con los programas escritos desde el proyecto para la recuperación de los datos o por el contrario la extracción debe hacerse de forma manual.

---

<sup>3</sup> Para mayor información véase <http://www.uib.no/en/project/polame>

En primera instancia, la elección de los periódicos se realizó con base en criterios de conocimiento contextual de los equipos de trabajo del proyecto, esto para garantizar que la información ofrecida por los mismos fuese relevante con relación a la política y economía de cada país (Chiquito & Quiroz, 2017, p.27).

Los criterios mínimos establecidos para la extracción de los artículos fueron:

- Título
- Sección
- Fecha
- URL (Uniform Resource Locator)
- Contenido
- Autor (opcional)

A continuación, se presenta la tabla 3 con la cantidad de artículos recuperados por periódico (Chiquito & Quiroz, 2017, p.35).

Tabla 3: *Cantidad de artículos recuperados por periódico*

<b>País</b>	<b>Periódico</b>	<b>Núm. artículos recuperados</b>
Argentina	Clarín	7570
	La Nación	3596
	La voz del interior	2458
	Página 12	1419
Brasil	Folha de Sao Paulo	7531
	O' Globo	11247
Colombia	El Tiempo	3782
	El Espectador	2442
México	El Universal	23763
	La Jornada	350

Los datos contenidos en la tabla anterior, corresponden a la información de los artículos recuperados por periódico de forma automática o semiautomática. En el caso de los periódicos no listados, la recuperación se realizó de forma manual y aleatoria hasta alcanzar el número de artículos necesarios para los propósitos del proyecto (Chiquito & Quiroz, 2017, p.34).

Para el caso específico del periódico El Colombiano, utilizado en el presente trabajo de investigación, el universo consta de 829 artículos. También fueron establecidos criterios para la extracción de la muestra de artículos para los propósitos del proyecto:

- ❖ Universo compuesto por los documentos descargados
- ❖ Años descargados entre 1990 y 2014 (según cada periódico)
- ❖ Más de 300 palabras y menos de 2000.
- ❖ Secciones relacionadas con la economía y la política
- ❖ Margen de error del 5 %
- ❖ Secciones Economía y Política homologadas
- ❖ +300 palabras y menos de 2000.

A continuación, se presenta la tabla 4 con el total de artículos por periódico de la muestra de trabajo (Chiquito & Quiroz, 2017, p.37).

Tabla 4: *Cantidad de artículos obtenidos en la muestra*

<b>País</b>	<b>Periódico</b>	<b>Artículos por muestra</b>
Argentina	Clarín	360
	La Nación	350
	La voz del interior	333
	Página 12	116
Brasil	Folha de Sao Paulo	367
	O’Globo	373
Colombia	El Tiempo	346
	El Espectador	332
México	El Universal	379
	La Jornada	350

La muestra de artículos para el periódico El Colombiano quedó conformada por un total de 312 artículos.

#### **4.4.2. DESCRIPCIÓN DE LA MUESTRA DE TRABAJO**

Se tomó una muestra de artículos del corpus PoLaMe de los tres periódicos más representativos en Colombia (El Colombiano, El Espectador y El Tiempo) de las secciones de política y economía.

La tabla 5 reúne la información del número total de artículos tanto en el Universo como en la muestra, y el número total de palabras por periódico en la muestra final (Chiquito & Quiroz, 2017, p.87).

Tabla 5: *Número de artículos y muestras de los periódicos*

	# de artículos (Universo)	# de artículo (muestra)	# de palabras (muestra)
<b>El Tiempo</b>	3782	346	206.553
<b>El Colombiano</b>	829	312	268.274
<b>El Espectador</b>	2442	332	17.730
<b>Totales</b>	7053	990	492.557

Para los objetivos del presente trabajo fueron seleccionados los artículos por medio de un muestreo aleatorio estratificado, teniendo como criterios mínimos de extracción: 1) que contuvieran la palabra pobreza; 2) que fuesen publicados entre los años 2010 y 2014; 3) que contaran con una extensión mínima de 300 palabras y máxima de 1000.

Nuestra muestra de trabajo quedó conformada por un total de 205 artículos (140.886 palabras), cuya distribución final se muestra en la figura 4: 56 artículos publicados en El Tiempo, 65 en El Espectador y 84 en El Colombiano.

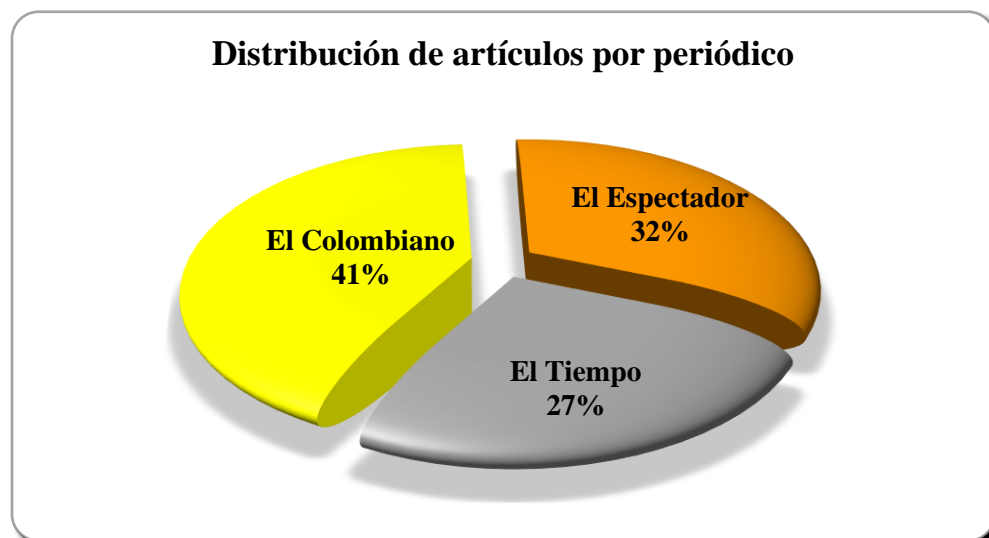


Figura 4: Distribución de artículos por periódico

## 4.5. HERRAMIENTAS Y RECURSOS

A través de las distintas etapas de la investigación, se ha hecho uso de distintas herramientas para cumplir con las tareas de procesamiento, almacenamiento y presentación de los datos. En primera instancia, se extrajeron las ocurrencias verbo más el sintagma nominal correspondiente al objeto directo. En esta etapa del proceso, se contó con la ayuda del ingeniero Antonio Jesús Tamayo Herrera, miembro del grupo de investigación en Traducción y Nuevas Tecnologías (TNT) perteneciente a la Escuela de idiomas de la Universidad de Antioquia, quien empleó distintas herramientas informáticas y diseñó los *scripts* en *Python* para la extracción de dichos elementos. Fueron extraídos un total de 6.283 ocurrencias del patrón [verbo + objeto directo] de los 205 artículos que constituyen la muestra de trabajo.

El programa empleado para el almacenamiento de los datos fue Excel 2016 de Microsoft, donde se cuenta con una hoja de cálculo con sus filtros respectivos para la presentación y manipulación de las ocurrencias. Otros programas como *EditPlus* 4.3 y *NoteTab Light* 7.2 (editores de texto), se utilizaron en las primeras pruebas con los datos para organizar la información y obtener algunos informes estadísticos descriptivos, como frecuencia de palabras, para posteriormente definir la submuestra de análisis y aplicar otras medidas estadísticas a nivel descriptivo.

Entre las diversas opciones ofrecidas por el mercado, se eligió el programa *StatGraphics* para el procesamiento estadístico de los datos. Se eligió Excel 2016 para la creación de tablas y gráficos en los análisis posteriores. Se tuvieron en cuenta cuatro elementos a la hora del marcaje de los datos. Por un lado, se realizó un marcaje semántico, donde se asignó tanto el tipo como el sentido a los verbos y se definió el área temática a la cual pertenecen cada uno de sus objetos directos. Por otro lado, se asignó la polaridad tanto al verbo como a su respectivo objeto directo. Finalmente, se estableció la carga global a cada patrón analizado.

Para el marcaje semántico del sentido de los verbos, se utilizó el programa *WordNet* 2.1 (marca registrada de la Universidad de Princeton). Se realizó un trabajo manual de traducción y verificación de cada verbo para asignar el sentido correspondiente de los propuestos en el programa. En cuanto al marcaje de las polaridades, para el caso de los verbos, se siguió el criterio lexicográfico de frecuencia en la organización de las acepciones. Este marcaje se realizó de manera

manual, puesto que se detectó un problema de equivalencia de significado entre el inglés y el español, principalmente en los objetos directos asignados en ambas lenguas. Esto se identificó, al realizar el mapeo y desambiguación de los verbos utilizando, entre otras, herramientas como *Freeling* y recursos como *SentiWordnet* en español.

Para el caso de los objetos directos, se realizó un marcaje manual para asignar su carga. También se contó con un ítem denominado *núcleo del objeto directo*, el cual se entiende como el sustantivo nuclear de un sintagma nominal presente dentro del objeto directo.

Para validar los datos, se contó con la colaboración de dos lingüistas expertos, miembros del grupo de Traducción y Nuevas Tecnologías TNT, de la Escuela de Idiomas de la Universidad de Antioquia; profesores Giovanni Diaz Laines y José Luis Rojas. Estos revisaron cada una de las ocurrencias (3.563) correspondientes al total de los datos luego de la depuración inicial y verificaron cada uno de los ítems empleados: sentido y tipo del verbo, carga del verbo, núcleo del objeto directo, carga del objeto directo y área temática. Al finalizar la validación de los datos, se revisaron nuevamente las ocurrencias, siguiendo las sugerencias e indicaciones de los validadores, quedando un total de 2.720, de las cuales se tomaron 1.360, correspondientes al 50 % de los datos, con las cuales se llevó a cabo el análisis.

#### **4.5.1. HERRAMIENTAS PARA EL ETIQUETAJE DE LOS DATOS**

Para hacer posible el análisis sintáctico y semántico de los datos, se utilizaron diversas herramientas o recursos. En primer lugar, se realizó el etiquetaje semántico de los verbos con el programa *WordNet 2.1* para determinar el sentido. En segundo lugar, el *Diccionario Combinatorio Práctico del Español Contemporáneo* dirigido por el Profesor Ignacio Bosque para determinar la carga de los verbos, al igual que el *Diccionario de la Real Academia de la Lengua Española* en sus versiones tanto impresas como digitales. A continuación, se presenta la descripción de estas herramientas y recursos.

##### **4.5.1.1. WORDNET 2.1**

Unido a lo descrito sobre la clasificación verbal en *WordNet*, el principal sistema de relaciones al interior de la base de datos es sinonimia, como el caso de *shut* y *close* (cerrar) o *car* (carro) y *automobile* (automóvil). Cada uno de los *synsets* están agrupados y conectados por una serie de



relaciones conceptuales, siendo intercambiables, en distintos contextos, aquellos que comparten el mismo concepto. También, se presenta una breve definición de cada uno y, en algunos casos, se ejemplifica su uso con frases cortas. De igual forma, cuando una ocurrencia tiene más de un significado, es presentada en tantos *synsets* como sea conveniente.

La estructura de organización corresponde básicamente a relaciones de super subordinación: hiperonimia e hiponimia, conectando *synsets* de uso general como el caso de *furniture*, *piece\_of\_furniture* (piezas del mobiliario) con aquellos de uso más específico *bed* (cama) y este a su vez con *bunkbed* (camarote). Otras relaciones de subordinación son establecidas, en el caso de los sustantivos, como la parte al todo, meronimia, en *chair* (silla) y *back*, *backrest* (espaldar). Es importante mencionar que en este tipo de subordinación no hay herencia de significado en las partes, es decir, una parte puede corresponder solo a un tipo específico de elementos y no al total de la clase.

#### **4.5.1.2. *DICCIONARIO COMBINATORIO PRÁCTICO DEL ESPAÑOL CONTEMPORANEO***

Como criterio de validación para el marcaje de la valencia de los verbos, se recurrió al *Diccionario Combinatorio Práctico del Español Contemporáneo* dirigido por el Profesor Ignacio Bosque. Se siguió el criterio de frecuencia en el orden de presentación en cada entrada de las acepciones, donde se anteponen aquellas de uso más frecuente y se posponen las de uso esporádico, criterio utilizado por la Real Academia Española en la presentación de la 22a edición del *Diccionario de la Real Academia de la Lengua Española*. La metodología que se siguió para este marcaje se describe con mayor especificidad en el apartado etiquetaje.

#### **4.5.1.3. *DICCIONARIO DE LA REAL ACADEMIA DE LA LENGUA ESPAÑOLA***

Al igual que el diccionario del Profesor Ignacio Bosque, se tomó como recurso para la asignación de las valencias de los verbos el *Diccionario de la lengua española* de la Real Academia Española en su versión digital, siguiendo el mismo criterio de orden en la indexación de las acepciones.

## 4.6. METODOLOGÍA

Para esta investigación, se ha establecido una metodología de trabajo general presentada en la figura 5. Estos son descritos con mayor especificidad en el desarrollo de cada capítulo. La metodología de trabajo establecida ha sido resumida en 5 etapas globales:

- a) *Corpus de trabajo PoLaMe*: ya se ha presentado una descripción del corpus que hemos tomado para llevar a cabo los análisis (Capítulo 4 § 4.5.1.).
- b) *Selección de la muestra para el análisis*: del total de artículos compilados en el corpus PoLaMe, se seleccionaron 205 artículos para conformar la muestra de trabajo.
- c) *Extracción de datos y validación*: teniendo como base la hipótesis que la polaridad de los textos recae en la relación [verbo + objeto directo], se crearon los *Scripts* en *Phyton* para extraer dichos datos, los cuales, fueron depurados y etiquetados posteriormente para ser validados con la ayuda de dos lingüistas expertos.
- d) *Estadística descriptiva*: a lo largo de la investigación, se aplicaron medidas estadísticas en diferentes momentos con la finalidad de establecer un seguimiento de la evolución de los datos y de este modo, presentar al final una descripción en detalle de los mismos.
- e) *Selección de muestra de datos para el análisis*: luego de finalizar la etapa de etiquetaje y depuración de los datos, se tomó la decisión de trabajar con el 50 % de estos, entre otras razones, por la cantidad de datos existente y el tiempo disponible para los análisis y la presentación final del escrito del trabajo de investigación. El 50 % de los datos analizados corresponden a la mitad de los datos totales y de allí que sea posible realizar la descripción de los patrones [verbo + objeto directo], unidades de análisis de nuestro trabajo de investigación.

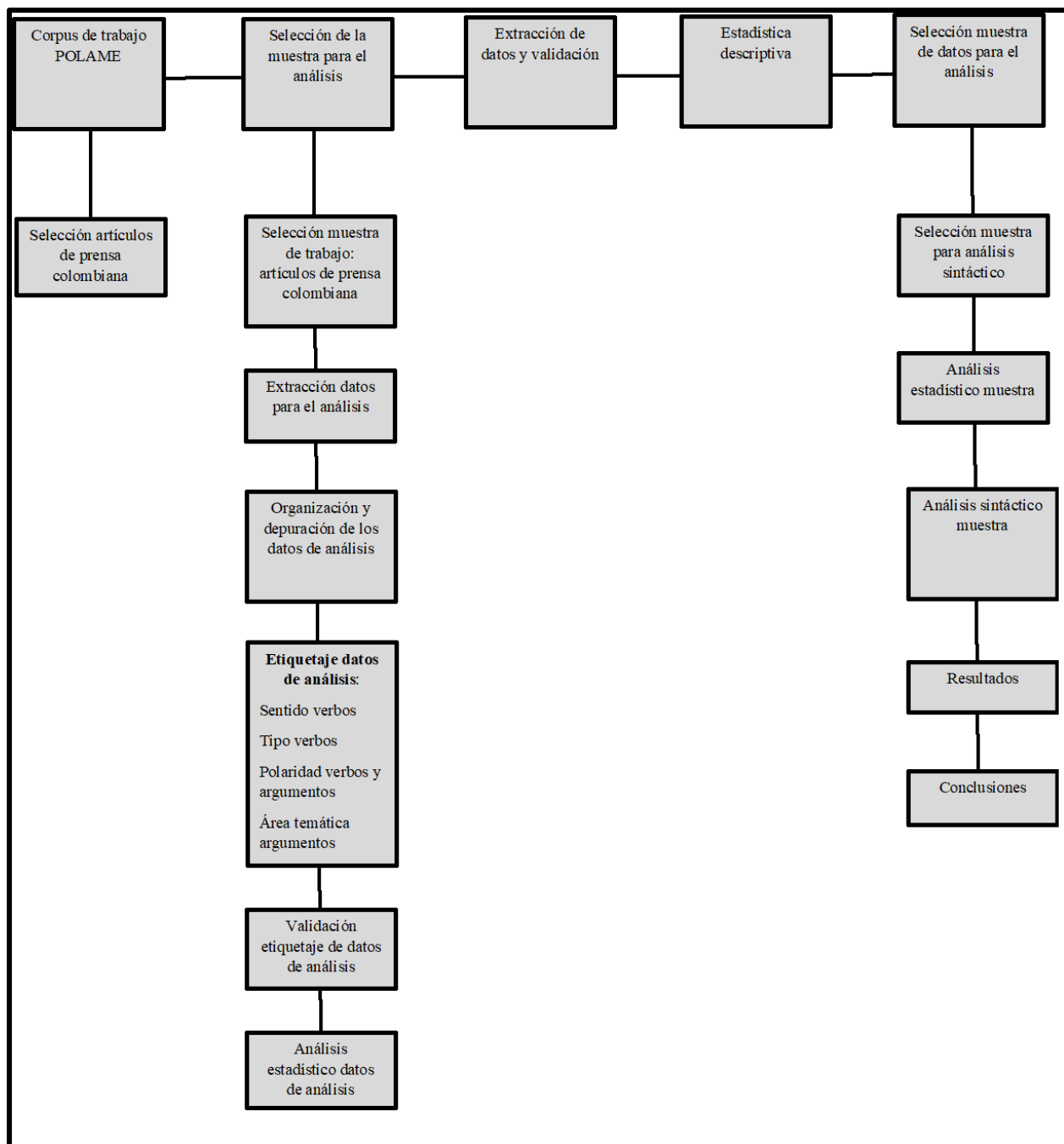


Figura 5: Esquema metodología de trabajo

#### 4.7. ETIQUETAJE

Con la finalidad de realizar una descripción en detalle de los datos, se procedió con el marcaje manual de cada una de las ocurrencias, con el fin de brindar una caracterización de cada una y de

este modo, lograr identificar las regularidades ofrecidas por los datos en la etapa de análisis y abordar las hipótesis planteadas. Se empleó Excel para la organización y posterior etiquetaje de los datos, como se muestra en la figura 6.

1	2	3	4	5	6	7	8	9	10	11	12	13	14
ID del artículo	URL	Periódico	Artículo completo	Negación	Verbo	Tipo de verbo	Sentido del verbo	Carga del verbo	Objeto Directo	Núcleo del Objeto Directo	Carga del objeto directo	Área temática	Carga patrón
1	http://www.elesp	El Espectador	pensionar más		disminuir	fraseológico	change	-1	la enorme desigualdad	enorme desigualdad	-1	economía	1
3	http://www.elesp	El Espectador	darle una		apoyar	fraseológico	stative	1	el proyecto europeo	proyecto europeo	0	política	1
7	http://www.elesp	El Espectador	asistencialismo	no	mejorar	fraseológico	change	1	la calidad de su	calidad	1	sociedad	-1
178	http://www.elcol	El Colombiano	soluciones para		aumentar	fraseológico	change	1	los niveles de	nivel de pobreza	-1	economía	-1

Figura 6: Nombres campos base de datos

Las columnas 1, 2, 3 corresponden a los metadatos; las columnas 4, 5, 6, y 10 corresponden a la información recuperada por medio de los *Scripts* en la etapa de extracción de los datos. Las columnas 7, 8, 9, 11, 12, 13 y 14 corresponden a la información que fue etiquetada. La base de datos puede dividirse en tres partes, de acuerdo con la información contenida.

En primer lugar, se presentan los metadatos: en la columna 1 (ID del artículo) se relaciona el número de identificación del artículo al interior de la base de datos. La columna 2 (URL) contiene la dirección *web* de donde fue extraído el artículo. En la columna 3 se relaciona el periódico en el cual fue publicado el artículo; y en la columna 4 se presenta el artículo completo.

En segundo lugar, se tiene la información extraída automáticamente para el análisis. En la columna 5 (negación) se relaciona la premodificación negativa presente en los patrones (43 ocurrencias). En la columna 6 están los verbos extraídos; mientras que en la columna 10 se encuentra el objeto directo que se une al verbo para formar los patrones de análisis [verbo + objeto directo].

En tercer lugar, se presenta la información obtenida en el proceso de etiquetado. En las columnas 7 y 8 (tipo y sentido del verbo, respectivamente) se relaciona la información sobre la tipología verbal empleada en nuestra investigación. La columna 11 corresponde al núcleo del objeto directo y en la columna 13 se relaciona el área temática del objeto directo. Las columnas 9, 12 y 14 se destinaron a las cargas del verbo, del objeto directo y del patrón de análisis, respectivamente. A continuación, se presenta la descripción de las etiquetas empleadas, teniendo en cuenta los criterios y/o los recursos utilizados en cada proceso.

#### 4.7.1. NÚCLEO DEL OBJETO DIRECTO

En primera instancia, se tomaron los objetos directos de la muestra y se identificó su núcleo, como puede verse en los ejemplos 52 y 53.

- |  |                       |
|--|-----------------------|
| 52. el embarazo adolescente                | embarazo adolescente. |
| 53. el sistema productivo del departamento | sistema productivo.   |

Por otro lado, hubo algunos casos en los cuales se dejaron los artículos, determinantes o preposiciones cuando no eran susceptibles de ser eliminados puesto que hacen parte del objeto directo como los casos 54 y 55.

- |                                 |                     |
|---------------------------------|---------------------|
| 54. un cierto nivel de activos  | nivel de activos    |
| 55. las brechas de la inequidad | brecha de inequidad |

#### 4.7.2. ÁREA TEMÁTICA DEL OBJETO DIRECTO

Las ocurrencias fueron etiquetadas de acuerdo con el área temática global correspondiente, integrando, bajo cada etiqueta, los términos que conforman un área de especialidad con los cuales se establecen entre sí relaciones basadas en las características que comparten (Marzá, 2012). De esta forma, cada ocurrencia fue etiquetada y posteriormente agrupada bajo un área temática común.

El criterio utilizado para definir la etiqueta de cada área temática corresponde a uno lexicográfico: cada etiqueta fue consultada en el *Diccionario de la Real Academia de la Lengua Española* y, de acuerdo a las acepciones allí listadas, se procedió a agrupar las distintas ocurrencias en 4 áreas temáticas, a saber: general, economía, política y sociedad.

A continuación, se presenta una breve descripción de cada etiqueta arriba mencionada. En la presentación de los resultados, se describen los componentes de cada una de las áreas temáticas y cómo estas fueron conformadas (Capítulo 6 § 6.1.9.).

#### **4.7.2.1. GENERAL**

Bajo esta etiqueta, se agruparon las ocurrencias ambiguas, es decir, las unidades sin rastro terminológico evidente. En el numeral 56, se recogen algunos ejemplos pertenecientes a esta categoría.

56. Potencia de locomotora, cambio de metodología, número aceptable, cobertura, participación, gran beneficio, consulta ágil, oportunidad de igualdad, mecanismo, impulso.

#### **4.7.2.2. ECONOMÍA**

Bajo esta etiqueta, se agruparon las ocurrencias relacionadas con la administración eficaz y razonable de los bienes, al igual que aquellas que dan cuenta de los bienes y actividades que integran la riqueza de una colectividad o individuo. En el numeral 57 se recogen algunos ejemplos pertenecientes a esta categoría.

57. Ingreso per cápita, dinero de Ecopetrol, reducción de costos, finanzas públicas, exportación de bienes, desempleo, precio medicamento, línea de pobreza, pobreza extrema, estrato.

#### **4.7.2.3. POLÍTICA**

Bajo esta etiqueta, se agruparon las ocurrencias correspondientes al arte, doctrina u opinión referente al gobierno de los Estados, al igual que aquellas relativas a quienes rigen o aspiran regir los asuntos públicos. En el numeral 58 se recogen algunos ejemplos pertenecientes a esta categoría.

58. Política del Alcalde, Gobierno del presidente Santos, gobierno colombiano, conflicto armado, reelección, línea de defensa de San Andrés, vicepresidencia, Fiscalía, justicia social, Agro Ingreso Seguro.

#### **4.7.2.4. SOCIEDAD**

Bajo esta etiqueta, se agruparon las ocurrencias correspondientes a todo lo relacionado con el conjunto de personas que conviven bajo normas comunes, es decir, elementos propios de la cultura,

el deporte, celebraciones, servicios, música, medios de transporte, religión, oficios, entre otros. En el numeral 59 se recogen algunos ejemplos pertenecientes a esta categoría.

59. Embarazo adolescente, componente de educación, salud, EPS, educación, reserva natural, consumo de sustancia psicoactiva, energía, lumbalú, calidad de vida.

#### 4.7.3. TIPO DE VERBO

Para establecer la tipología verbal, se retomó la propuesta presentada por Lorente (2002) en su trabajo sobre los verbos en el discurso especializado. De los cuatro tipos de verbos en Lorente, hemos encontrado en los datos, verbos de tipo fraseológico, discursivo y conector. La figura 7, presenta la relación entre los verbos y su tipología en la muestra de trabajo.

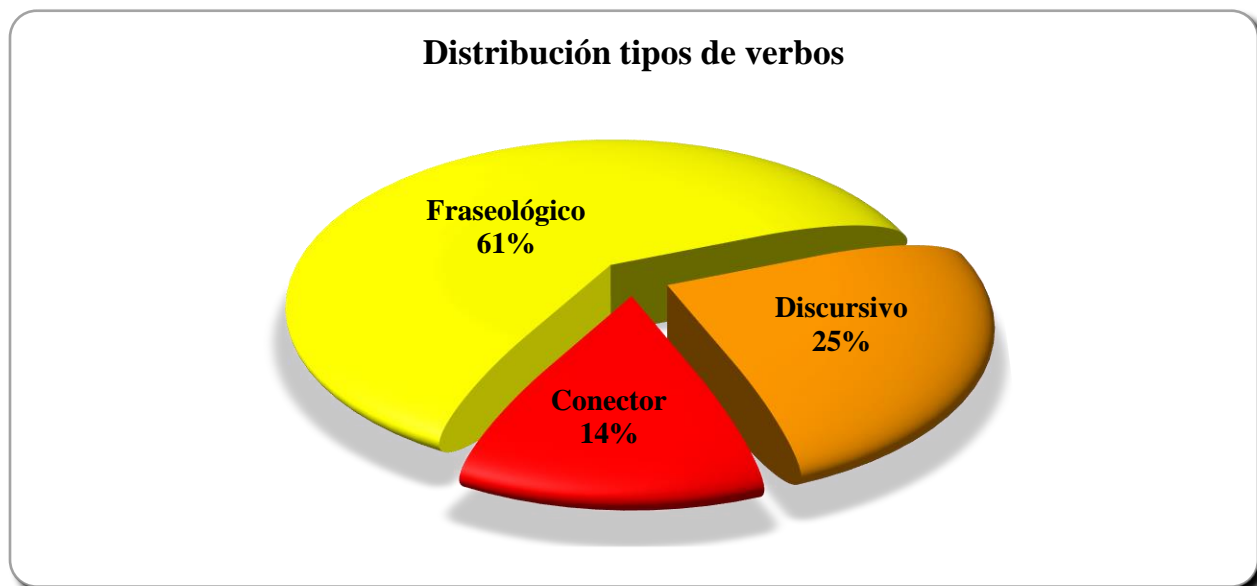


Figura 7: Distribución tipos de verbos

#### 4.7.4. SENTIDO DEL VERBO

Para asignar los sentidos a los verbos, se trabajó con la base de datos *WordNet 2.1*, descrita en el numeral sobre las herramientas para el etiquetaje de los datos. La asignación se realizó manualmente. En primer lugar, se buscó el equivalente en lengua inglesa del verbo en español, para ello se hizo uso del diccionario *Wordreference*. En segundo lugar, se recuperó cada entrada

en *WordNet* 2.1. y se seleccionó el sentido apropiado para cada una. La figura 8, ilustra la interfaz de búsqueda de *WordNet* 2.1.

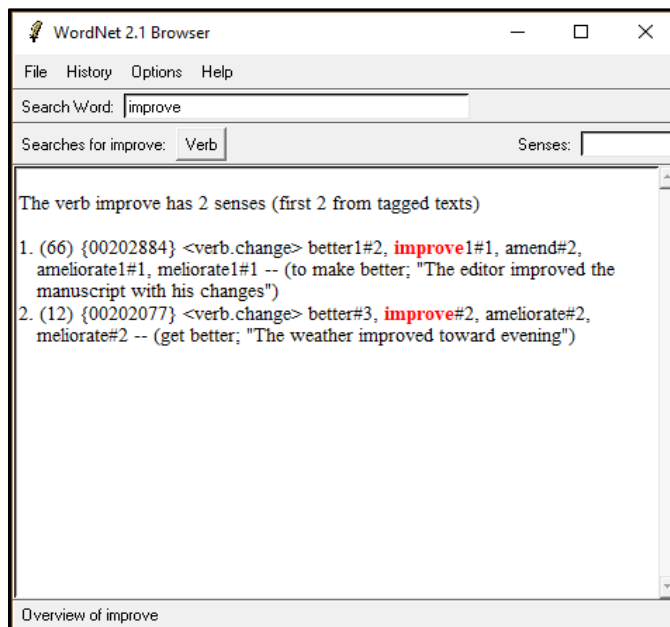


Figura 8: Interfaz *WordNet* 2.1

#### 4.7.5. CARGAS DE LOS VERBOS Y DE LOS OBJETOS DIRECTOS

La asignación de las cargas se realizó de manera manual, siguiendo de nuevo el criterio de frecuencia de las acepciones ofrecidas por el *Diccionario Combinatorio Práctico del Español Contemporáneo* del profesor de Monte y/o el *Diccionario de la Real Academia de la Lengua Española*. De esta forma, se limitó la asignación de las cargas, contando con las ocurrencias más frecuentes.

#### 4.8. ESTADÍSTICA

La estadística de los datos, se realizó con la finalidad de presentar una descripción de las ocurrencias y realizar los análisis correspondientes. Puntualmente, se llevó a cabo un análisis descriptivo de los datos, tales como frecuencias de aparición y algunos cruces de variables partiendo de la información arrojada por las tablas de frecuencia. Los datos son presentados ya sea en tablas o gráficas circulares o de barras. La información resultante de los análisis estadísticos es presentada en la Parte IV: Análisis de datos y resultados (Capítulo 6).



## **CAPÍTULO 5: METODOLOGÍA PARA LA CONSTRUCCIÓN DE UNA BASE DE DATOS [VERBO + OBJETO DIRECTO]**

En el capítulo anterior, se presentó el paso a paso en el análisis de los patrones [verbo + objeto directo], tenidos en cuenta en la construcción de una base de datos, con miras a aportar soluciones y mejorar los resultados de las herramientas diseñadas para el AS en artículos de prensa colombiana. Teniendo claro que la propuesta metodológica acá presentada no es una solución al problema del AS, consideramos que esta podría constituirse en un aporte significativo a la consecución de la misma. El objetivo del presente capítulo busca presentar, en forma de etapas, los tres momentos en la construcción de la base de datos: extracción de las ocurrencias del patrón, depuración de los datos y etiquetado.

### **5.1. ETAPA 1: EXTRACCIÓN DE PATRONES [VERBO + OBJETO DIRECTO]**

La extracción de los patrones de análisis correspondientes al predicado de las oraciones, fue realizada con ayuda del ingeniero Antonio Jesús Tamayo Herrera. En la extracción se tuvieron en cuenta tres tipos de patrones poliléxicos: patrones con el atributo de los verbos copulativos o pseudocopulativos; patrones [verbo + objeto directo]; y patrones con adjetivos y complementos adjetivales.

Para los objetivos de nuestro trabajo de investigación, se limitó el trabajo a los patrones [verbo + objeto directo]. Hemos considerado que cada una de los patrones mencionados, merece un trato especial; es decir, cada uno debería ser abordado individualmente en futuros trabajos de investigación, dado desbordaría los límites del tipo de trabajo acá realizado, especialmente, en términos de tiempo. Los patrones para nuestro análisis fueron extraídos por medio de un *Script* en *Python*, diseñado para recuperar la información correspondiente a dicho patrón. En el numeral 60 se muestran ejemplos de las ocurrencias extraídas de este patrón.

60. abordar aquellas trampas, disminuir los niveles de pobreza, generar empleo, implementar programa social, incrementar presupuesto, mejorar calidad de vida, reducir índice de pobreza extrema, superar carencia básica, trabajar inclusión social.

## 5.2. ETAPA 2: DEPURACIÓN DE LAS OCURRENCIAS EXTRAÍDAS

En la segunda etapa, se procedió con la depuración de las ocurrencias extraídas en la etapa 1. Se llevó a cabo un análisis inicial de los datos para definir los criterios de depuración, descritos a continuación.

### 5.2.1 ELIMINACIÓN DE ERRORES EN LA EXTRACCIÓN

Teniendo presente el porcentaje de error en la extracción de los datos por medio del *Script*, se eliminaron las ocurrencias que extraídas y que no corresponden a la información solicitada. La tabla 6, contiene algunas de las ocurrencias eliminadas identificadas como errores de extracción.

Tabla 6: Ejemplos de ocurrencias eliminadas: errores de extracción

<b>Id del artículo</b>	<b>Negación</b>	<b>Verbo</b>	<b>Objeto Directo</b>
98		given	all
112		holandésel	programa
124		niñoshay	19 sentencias condenatorias
126		however	promising
138		gratisramiro	Velásquez_Gómez

### 5.2.2 ELIMINACIÓN DE PRONOMBRES

Se eliminaron las ocurrencias correspondientes a los pronombres de tipo personal, indefinido, relativos, reflexivos, interrogativos. La tabla 7, contiene ejemplos de algunas de las ocurrencias eliminadas según los pronombres listados.

Tabla 7: Ejemplos de ocurrencias eliminadas: pronombres

<b>Id del artículo</b>	<b>Negación</b>	<b>Verbo</b>	<b>Objeto Directo</b>
11		prevenir	esto
12		imaginar	usted
12	no	saber	lo que
62	no	saber	quién
62		pensar	yo mismo
87		acceder	que
131		parar	nosotros
137		corregir	algunos
139		creer	nada

145		afirmar	esto
154	no	saber	qué
155	no	suceder	lo mismo
164		capacitar	él
174		establecer	aquellas

### 5.2.3 ELIMINACIÓN DE DEÍCTICOS

Las ocurrencias correspondientes al uso de deícticos en el discurso, fueron también eliminadas de la base de datos, como se muestra en la tabla 8.

*Tabla 8:* Ejemplos de ocurrencias eliminadas: deícticos

<b>Id del artículo</b>	<b>Negación</b>	<b>Verbo</b>	<b>Objeto Directo</b>
110		acercar	los presentes
113		priorizar	muchas de sus acciones
141		acometer	el Gobierno
156		hacer	las veces de enfermera
200		aclarar	Gómez

### 5.2.4 ELIMINACIÓN DE FORMAS IMPERSONALES O INCOMPLETAS

Las formas impersonales o las formas incompletas se eliminaron de la base de datos, puesto no aportan ningún tipo de información relevante con relación a las ocurrencias recuperadas. La tabla 9, contiene algunos ejemplos de las ocurrencias eliminadas correspondientes a las formas impersonales o incompletas.

*Tabla 9:* Ejemplos de ocurrencias eliminadas: formas impersonales o incompletas

<b>Id del artículo</b>	<b>Negación</b>	<b>Verbo</b>	<b>Objeto Directo</b>
87		ubicar	Los
93		utilizar	Sin
98		hacer	algo
122		calificar	uno
159		soñar	algo

### 5.2.5 ELIMINACIÓN DE OCURRENCIAS REPETIDAS

En algunos casos, se encontró una ocurrencia en varias ocasiones. Las ocurrencias repetidas se eliminaron de la base de datos dejando únicamente una ocurrencia, como se muestra en la tabla 10.

Tabla 10: Ejemplos de ocurrencias repetidas eliminadas

<b>Id del artículo</b>	<b>Negación</b>	<b>Verbo</b>	<b>Objeto Directo</b>
62		decir	eso
62		decir	eso
62		decir	eso
65		comprar	esa leche
65		comprar	esa leche
65		dar	instrucciones
65		dar	instrucciones

### 5.2.6 ELIMINACIÓN DE CIFRAS

Con relación a las cifras, se tomó la decisión de eliminarlas por cuestiones metodológicas. El tratamiento de las ocurrencias con números o referencia clara a cifras ameritan, sin lugar a dudas, un tratamiento especial e incluso podrían ser objeto de estudio en un nuevo trabajo de investigación. La tabla 11, contiene algunos ejemplos de ocurrencias de este tipo.

Tabla 11: Ejemplos de ocurrencias con cifras

<b>Id del artículo</b>	<b>Negación</b>	<b>Verbo</b>	<b>Objeto Directo</b>
32		medir	0
51		beneficiar	1.151.000_hectáreas
51		recibir	27.371 créditos
54		pasar	\$19,3 billones
107		superar	16_mil_millones_de_dólares

Sin embargo, algunas ideas fueron discutidas con relación a las ocurrencias con cifras y se obtuvieron algunas conclusiones como se muestra en la tabla 12. En los numerales 61 a 64, se presentan algunos ejemplos con relación a estas conclusiones.

Tabla 12: Reglas con ocurrencias que contienen cifras

<b>Carga del verbo</b>	<b>Cifra</b>	<b>Carga del objeto directo</b>	<b>Polaridad final del patrón</b>
------------------------	--------------	---------------------------------	-----------------------------------

Positivo		Positivo	Positivo
Positivo		Negativo	Negativo
Negativo		Positivo	Negativo
Negativo		Negativo	Positivo

61. Invertir \$230 billones

62. Lograr 10 mil cupos nuevos

63. Perder 20 millones de pesos

64. Construir 3.500 viviendas

### 5.2.7 ELIMINACIÓN DE NOMBRES PROPIOS

Otro elemento que no fue tenido en cuenta en nuestro análisis, corresponde a las ocurrencias que contenían nombres propios. La decisión de omitir este tipo de información obedece a que, en términos de polaridad, la función del nombre más que aportar carga (positiva, negativa o neutra) está enfocada a otro tipo de información complementaria; es decir, información sobre quién o en quién recae la transitividad del verbo o la información relativa a lugares. La tabla 13, contiene algunos ejemplos de ocurrencias con nombres propios eliminadas durante el proceso de depuración del corpus.

Tabla 13: Ejemplos de ocurrencias con nombres propios

Id del artículo	Negación	Verbo	Objeto Directo
39		explicar	Sanabria
60		expresar	Emilio_García_Méndez
85		decir	el presidente Santos
87		agregar	Zavala
144		entrevistar	Ana Patricia Aristizábal

### 5.2.8 ELIMINACIÓN DE LOS VERBOS MODALES

Según Gaya (1980), los verbos modales expresan el *modus* explícito de la oración, mientras el verbo en infinitivo que los acompaña expresa el *dictum*. Debido al carácter auxiliar de los verbos modales, hemos tomado la decisión de excluirlos de nuestro análisis, toda vez que las características inherentes a estos contribuyen a especificar el significado del verbo que auxilian.

Por lo tanto, no cumplen con el patrón de análisis que hemos establecido [verbo + objeto directo] debido a que estos siempre acompañan a otro verbo.

### 5.3. ETAPA 3: ETIQUETADO DE LAS OCURRENCIAS DEPURADAS

Un elemento importante en nuestro trabajo, hace relación al análisis sintáctico de las ocurrencias, dado que permite la identificación de los distintos componentes de cada una, al igual que sus funciones y relaciones al interior de las mismas. En la etapa de extracción, se utilizó la herramienta *Freeling* para recuperar la información sintáctica de los textos analizados y posteriormente, diseñar el *Script* para la extracción del patrón de análisis de la investigación. En la figura 9 se presenta un ejemplo de análisis sintáctico realizado con el programa mencionado.

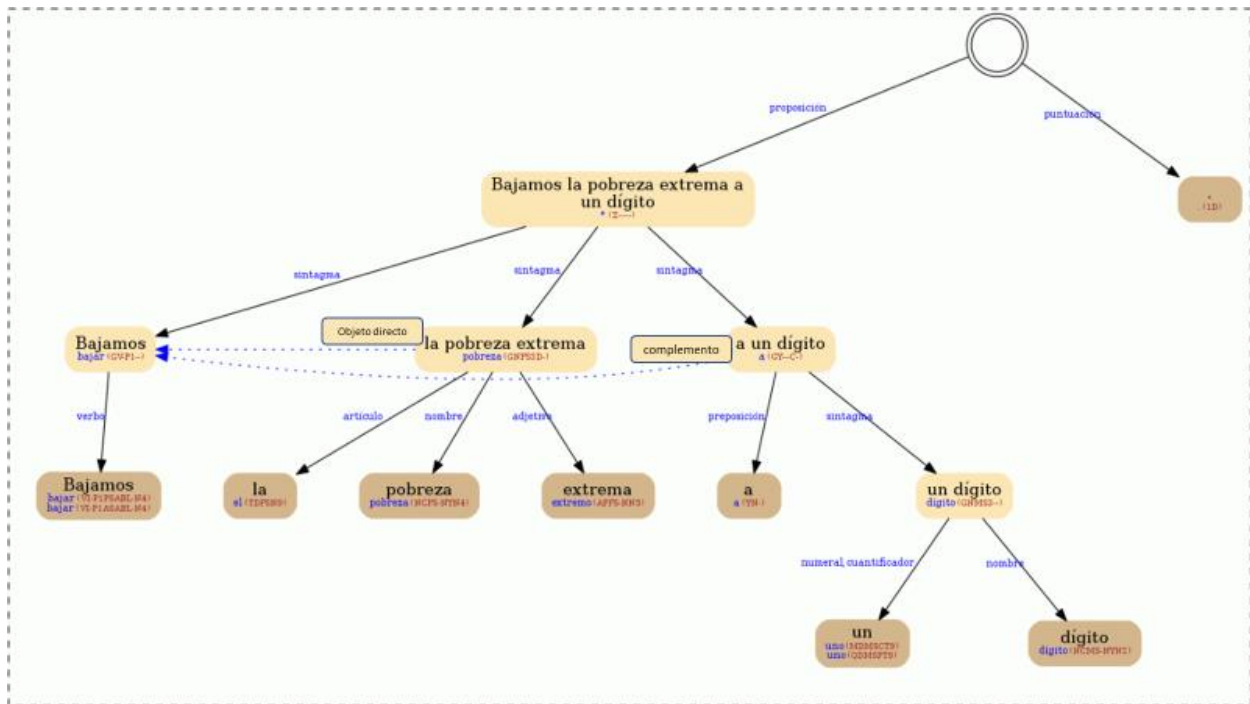


Figura 9: Ejemplo árbol sintáctico

En la etapa 3, se procedió con el etiquetaje semántico de las ocurrencias, al igual que con la asignación de las cargas de polaridad. Se etiquetó tanto el tipo como el sentido a cada verbo y se definió el área temática de los objetos directos. Con relación a las cargas de polaridad asignadas a las ocurrencias, se tomó en cuenta el verbo y el objeto directo individual. Al final se asignó una carga total al patrón de análisis.

### 5.3.1. ANÁLISIS SINTÁCTICO

Con relación al aspecto sintáctico de las ocurrencias, se tomó el verbo como núcleo de la oración para posteriormente recuperar la información correspondiente a las negaciones del tipo *no* y *no se* cuándo estas aparecen como premodificadores. Estas las hemos denominado como *conmutadores* de polaridad, puesto que en las ocurrencias donde se encuentran, su función es la de *conmutar* la carga final asignada al patrón de estudio. Además, se recuperó el sintagma nominal adyacente al verbo que corresponde al objeto directo. La tabla 14, contiene las ocurrencias obtenidas por medio de los *Scripts* de extracción, tal como aparecen en el archivo inicial (antes de ser procesadas en el archivo final).

Tabla 14: Ejemplo ocurrencias extraídas

<b>Id de artículo</b>	<b>Negación</b>	<b>Verbo</b>	<b>Objeto Directo</b>
1		alcanzar	este anhelado propósito
2		examinar	los avances
4	no se	reconocer	la crisis de la salud
4	no	cotizar	pensiones
200		incorporar	las transferencias monetarias de el Gobierno
203		seleccionar	un grupo de 4.000 trabajadores caracterizados

### 5.3.2. ETIQUETAJE SEMÁNTICO

Con el etiquetaje semántico, se pretendió caracterizar no solo el verbo en cuanto al tipo y el sentido, sino también, identificar los temas en los cuales se enmarcan los artículos analizados, por medio del etiquetaje de los objetos directos de acuerdo a sus áreas temáticas. A Continuación, se presentan algunos ejemplos del etiquetaje semántico de las ocurrencias.

#### 5.3.2.1. TIPO DE VERBO

Siguiendo la propuesta de Lorente (2002) sobre la tipología de los verbos en el discurso especializado, se clasificaron los verbos, de acuerdo al tipo correspondiente, como se muestra en la tabla 15.

Tabla 15: Ejemplos etiquetaje del tipo de verbo

<b>Id del artículo</b>	<b>Negación</b>	<b>Verbo</b>	<b>Tipo de verbo</b>
62		afectar	Conector

277		entregar	Conector
293		explicar	Discursivo
120		destacar	Discursivo
115		aumentar	Fraseológico
149		reducir	Fraseológico

### 5.3.2.2. SENTIDO DE VERBO

Para el sentido de los verbos, se tomó como guía la propuesta del proyecto *WordNet*, según la cual, existen 15 categorías de sentidos de verbos agrupados jerárquicamente de acuerdo con las relaciones de troponimia, tal como se explicó en el marco teórico (Capítulo 3 § 3.1.1.4.1.2.). La tabla 16, reúne algunas ocurrencias con relación al etiquetaje del sentido de los verbos.

Tabla 16: Ejemplos etiquetaje del sentido de verbo

<b>Id del artículo</b>	<b>Negación</b>	<b>Verbo</b>	<b>Sentido de verbo</b>
13		afectar	<i>Change</i>
172		entregar	<i>Possession</i>
232	no se	preocupar	<i>Emotion</i>
5		destacar	<i>Communication</i>
134		presentar	<i>Perception</i>
225		apoyar	<i>Stative</i>

### 5.3.2.3. ÁREA TEMÁTICA

Por último, se asignó el área temática a cada uno de los objetos directos. Estos se clasificaron teniendo en cuenta los campos temáticos comunes, es decir, bajo un área temática se reúnen las ocurrencias que comparten un mismo núcleo temático, tal como se explicó en la metodología (Capítulo 4 § 4.7.2) y se especificó en los resultados (Capítulo 6 § 6.1.9). La tabla 17, da cuenta del etiquetaje de los objetos directos conforme al área temática.

Tabla 17: Ejemplos etiquetaje del área temática del objeto directo

<b>Id del artículo</b>	<b>Verbo</b>	<b>Objeto directo</b>	<b>Área temática</b>
146	cumplir	los Objetivos de desarrollo del Milenio	Economía
221	dar	tranquilidad	General
99	seguir	espacio político	Política
212	entregar	sistema de salud	Sociedad



## 5.4. ASIGNACIÓN DE LAS CARGAS DE POLARIDAD

Para la asignación de las cargas de polaridad, se tomaron dos elementos: el verbo y el objeto directo. Al final, se realizó la suma de las cargas del verbo y del objeto directo para obtener la polaridad final del patrón analizado. Se tuvo en cuenta la premodificación negativa de algunas ocurrencias a la hora de asignar la carga al patrón final.

Las cargas asignadas son: positiva (identificada en la base datos con el número 1), negativa (identificada con -1), y neutra (identificada con 0). Se establecieron variables discretas para la asignación de la polaridad. La tabla 18, da cuenta de la asignación de las cargas de polaridad a los verbos y a los objetos directos, al igual que la carga del patrón final.

*Tabla 18:* Ejemplos etiquetaje de la polaridad de los elementos de las ocurrencias

<b>Id del artículo</b>	<b>Verbo</b>	<b>Carga verbo</b>	<b>Objeto directo</b>	<b>Carga objeto directo</b>	<b>Carga patrón</b>
145	generar	positivo	mensaje equivocado	negativo	Negativo
146	lograr	positivo	seguridad	positivo	Positivo
82	combatir	negativo	narcotráfico	negativo	Positivo
188	sufrir	negativo	embate de la guerra	negativo	Negativo
29	llegar	neutro	paz total	positivo	Positivo
168	hacer	neutro	estudio reciente	neutro	Neutro

Al finalizar el proceso de etiquetaje, se procedió con el análisis de la información contenida en la base de datos depurada, la cual se presenta en la Parte IV del presente trabajo de investigación.

# **PARTE IV: ANÁLISIS DE DATOS Y RESULTADOS**



## CAPÍTULO 6: ANÁLISIS DE DATOS Y RESULTADOS

Luego de una depuración sistemática de los datos donde, paso a paso, se fue eliminando la información no relevante para los objetivos de la investigación y, teniendo en cuenta que los datos eliminados de nuestra base de datos son susceptibles de ser analizados en investigaciones futuras, para enriquecer el campo de trabajo; el presente capítulo contiene los resultados del análisis realizado con los patrones [verbo + objeto directo] correspondientes al 50 % (1.360 ocurrencias) de los datos finales. Es importante precisar que, cuando se utiliza la palabra *ocurrencia*, se hace referencia al total de elementos que componen la base de datos, los cuales corresponden al *patrón* del análisis, entendido como la estructura [verbo + objeto directo].

Se dividió la presentación de los resultados obtenidos en la investigación en dos partes. En primer lugar, se presentan las tablas de frecuencia junto a la descripción de cada una de ellas, enfatizando en las categorías utilizadas en el análisis. Los datos de cada tabla son organizados de forma descendente, partiendo desde el de mayor frecuencia hasta llegar al de menor. Es importante precisar que, aunque se trabajó con el 100 % de los datos, al momento de presentar los resultados de nuestro análisis, se tomaron en cuenta solo los valores por encima del 5 %. Así mismo, se presentaron anotaciones sobre los valores inferiores al punto de corte establecido para precisar los análisis.

En segundo lugar, junto con la descripción de los resultados, se presentó una interpretación de los datos, no solo en términos de las polaridades asignadas a cada componente de la ocurrencia analizada, sino también con relación a la información ofrecida por estas. Este análisis obedece a las posibilidades que ofrece el AS en la interpretación de los fenómenos estudiados (la pobreza en nuestro caso) y en la construcción de un acercamiento teórico y metodológico de los referentes temáticos trabajados. Adicional a las descripciones mencionadas, se puntualizaron cada uno de los elementos que ayudan a resolver tanto la hipótesis como los objetivos planteados en la investigación.

## **6.1. ANÁLISIS DE DATOS Y RESULTADOS**

La pobreza como fenómeno social ha existido a la par con el ser humano. Culturalmente, las sociedades han establecido distintas formas de jerarquía donde siempre debe haber alguien en el fondo (Øyen, 2004). Este estudio sobre la pobreza en la prensa colombiana, desde la construcción de una base de datos con fines de AS, nos ha permitido caracterizar las unidades lingüísticas conformadas por los patrones [verbo + objeto directo], identificando los de mayor discriminación con relación a la polaridad de los enunciados.

De igual modo, la interpretación de estos patrones permitiría establecer los referenciales mediáticos de la pobreza y el bienestar en Colombia en la forma como es visto el problema, puesto que “la acción mediática y sus efectos sobre la construcción de los referenciales de acción pública tienen como punto de referencia la actividad discursiva”(Pardo & Ruiz, 2017).

El análisis retoma únicamente aspectos verbales dejando a un lado otros elementos igualmente importantes en la construcción de un concepto de pobreza en Colombia (pertenecientes al campo de la semiótica ajeno a nuestro análisis). Los resultados aquí presentados, constituyen un avance en la metodología para la construcción de una base de datos lingüística con el verbo como unidad de análisis, unido a sus objetos directos con propósitos de AS. Así mismo, permitiría un acercamiento a la forma como se percibe el fenómeno de la pobreza en la prensa colombiana a partir del análisis de estas ocurrencias.

Los análisis y las conclusiones que se describen en el presente trabajo de investigación, tienen una validación interna y no establecen una generalización sobre el fenómeno abordado; teniendo en cuenta que no se realizaron pruebas en sistemas de AS, con las cuales se podría lograr una verificación externa de los datos.

### **6.1.1. DEL UNIVERSO A LA MUESTRA DE TRABAJO**

Para los fines de la presente investigación, se tomaron los periódicos colombianos recopilados en el proyecto PoLaMe: El Colombiano (diario de distribución regional), El Espectador y El Tiempo (de distribución nacional). El Universo de datos corresponde a un total de 7.053 artículos

compilados de forma automática, semiautomática o manual; de los cuales, 990 corresponden a la muestra de trabajo en PoLaMe.

Se seleccionó un total de 205 artículos para conformar la muestra de trabajo de esta investigación, de los cuales se extrajeron las ocurrencias de análisis. La tabla 19, adaptada de la presentada en Quiroz, Chiquito, & Zuluaga (2017), resume esta información que va desde el universo de datos hasta nuestra muestra de trabajo.

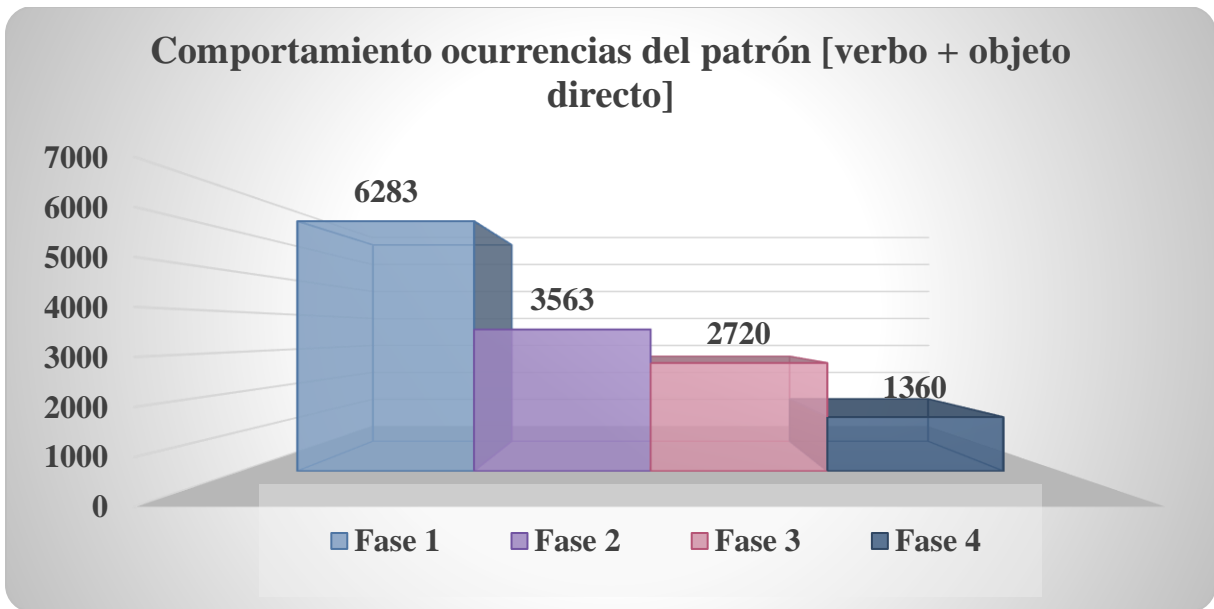
*Tabla 19:* del universo de datos a la muestra de trabajo

	# de artículos (universo PoLaMe)	# de artículo (muestra PoLaMe)	# de palabras (muestra PoLaMe)	# de artículo (muestra trabajo)	# de palabras (muestra trabajo)
<b>El Tiempo</b>	3782	346	206.553	56	30239
<b>El Colombiano</b>	829	312	268.274	65	61722
<b>El Espectador</b>	2442	332	17.730	84	47012
<b>Totales</b>	7053	990	492.557	205	138973

Con relación a los patrones [verbo + objeto directo] que se recuperaron de la muestra de trabajo, podemos establecer cuatro fases en el comportamiento de los mismos:

- ❖ Fase 1: aquí se relaciona el 100 % de las ocurrencias tal y como se extrajeron a partir del *Script*. Un total de 6.283 ocurrencias dan cuenta de esta primera fase.
- ❖ Fase 2: esta hace referencia a la cantidad de ocurrencias, luego de llevar a cabo la depuración. En esta etapa, quedaron un total de 3.563 ocurrencias.
- ❖ Fase 3: este momento es posterior a la validación de los datos. Luego de las recomendaciones de los expertos, se eliminaron 843 ocurrencias, quedando un total de 2.720 que conforman los datos finales de la investigación.
- ❖ Fase 4: por último, se tomó el 50 % de los datos resultantes en la fase 3, con los cuales se llevó a cabo el análisis. La decisión de trabajar solo con el 50 % de los datos, fue de carácter metodológico, con relación a la delimitación del tiempo.

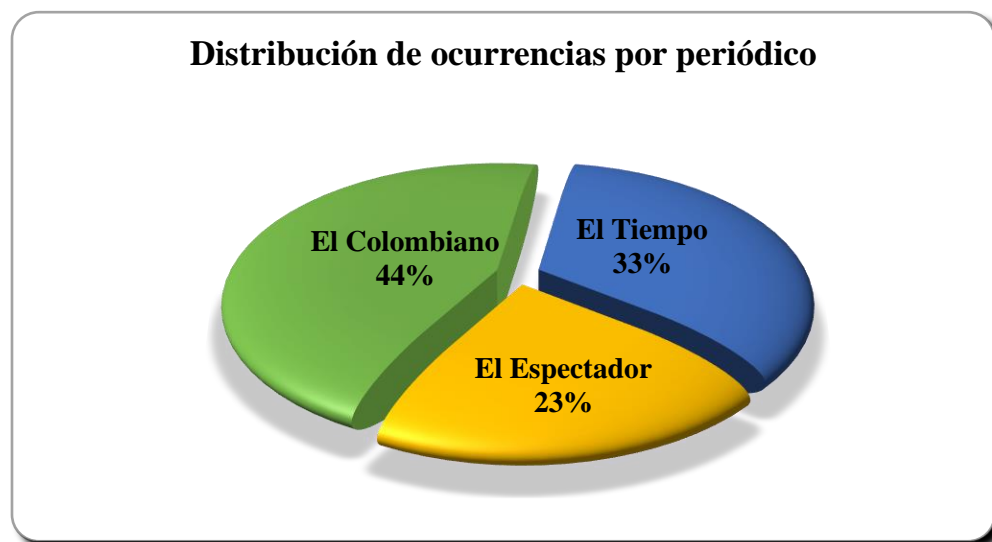
La figura 10, da cuenta del comportamiento de las ocurrencias desde la extracción inicial hasta la conformación de la muestra de trabajo.



*Figura 10:* Comportamiento ocurrencias del patrón [verbo + objeto directo]

### 6.1.2. DISTRIBUCIÓN DE LAS OCURRENCIAS POR PERIÓDICO

Después de la extracción, depuración y posterior selección de la muestra, las 1.360 ocurrencias analizadas quedaron distribuidas de la siguiente manera: en el periódico El Colombiano, se registraron 600 ocurrencias (44,12 %); 448 ocurrencias en El Tiempo (32,94 %); y 312 ocurrencias (22,94 %) en El Espectador, como se muestra en la figura 11.



*Figura 11:* Distribución de ocurrencias por periódico

En cuanto a la distribución de las cargas por periódico, en los tres casos la frecuencia más alta corresponde a los patrones marcados como positivos, mientras los negativos ocupan el segundo lugar en El Colombiano y el tercer lugar en el Tiempo y el Espectador. En cuanto a los neutros, ocupan el tercer lugar en El Colombiano y el segundo lugar en El Tiempo y El Espectador. En la tabla 20, se resume la información sobre la distribución por periódico de las ocurrencias de acuerdo a la carga final asignada.

Tabla 20: Distribución cargas patrón por periódico

Carga	El Colombiano		El Espectador		El Tiempo	
	Frecuencia	%	Frecuencia	%	Frecuencia	%
Positivo	387	64,5	301	67,19	202	64,74
Negativo	122	20,33	74	16,52	48	15,38
Neutro	91	15,17	73	16,29	62	19,87
Palabras o elementos distintos contados:	3	100	3	100	3	100
Total:	600		448		312	

En la literatura del AS, la definición de las categorías de polaridad se da por sobreentendida en el ámbito académico. Sin embargo, el neutro (a diferencia de lo positivo y negativo) al presentar un mayor desafío en la construcción de herramientas para el AS, ha sido definido como “no opinión” (Liu, 2015, p.9) o “*lack of opinion*” (Pang & Lee, 2008, p.27).

Así mismo, para puntualizar aquello que se entiende por positivo o negativo, hacemos uso de la definición de estos términos en el *Diccionario de la Real Academia* y a su vez los contrastamos con el *Diccionario Combinatorio Práctico del Español Contemporáneo* de Ignacio Bosque y el *Diccionario de Uso del Español* de María Moliner.

Según el *Diccionario de la Real Academia*, se entiende como positivo al adjetivo que es “Cierto, efectivo, verdadero y que no ofrece duda” (Real Academia Española, 2017). Por otro lado, en el *Diccionario Combinatorio Práctico del Español Contemporáneo*, positivo es igualmente un adjetivo y se define como algo “afirmativo o a favor” (Bosque, 2006). Por último, en el *Diccionario de Uso del Español* de María Moliner, lo positivo también es definido como un adjetivo que:

se aplica, por oposición a negativo, a un valor, un estado, una cualidad, etc., que consiste en la existencia y no en la falta de algo: ‘saldo positivo de una cuenta. Una prueba positiva. Es bueno, no por falta de maldad, sino con bondad positiva’ (Moliner, 2007).

Por otro lado, con relación a lo negativo, este es presentado en el *Diccionario de la Real Academia* como un adjetivo "perteneciente o relativo a la negación" en su primera acepción, que no ofrece mayor claridad, por lo que tomamos la cuarta acepción donde se expresa mejor la definición de lo negativo como adjetivo en cuanto hace referencia a “una cosa: Mala, perjudicial o infructuosa” (Real Academia Española, 2017). En el *Diccionario Combinatorio Práctico del Español Contemporáneo*, lo negativo, como adjetivo, se entiende como algo “desfavorable o perjudicial” (Bosque, 2006). En el *Diccionario de Uso del Español*, lo negativo es también un adjetivo y “se aplica a lo que contiene una negación o sirve para negar: ‘una respuesta negativa. Forma negativa del verbo’. No se dice *adverbio de negativo*, sino *de negación*” (Moliner, 2007).

Teniendo en cuenta que el eje temático de la investigación es la pobreza, hemos de entender lo positivo como aquello que favorece la erradicación de esta condición o el mejoramiento de los criterios con la cual se mide. Así mismo, lo negativo se entiende como aquello que permite que el fenómeno se aumente o se mantenga; es decir, que haya más pobreza. Con relación al neutro, se presenta una definición teniendo como base los resultados de la investigación (Capítulo 6 § 6.1.5.3). En esta perspectiva, siendo la pobreza una temática en sí misma negativa, se puede observar como el comportamiento de la polaridad de los patrones, podría evidenciar la forma cómo es abordado el fenómeno en los periódicos; los cuales, basados en lo propuesto por Chomsky (1997), establecen la agenda política del país.

### **6.1.3. FRECUENCIA DE LOS VERBOS**

Un total de 89 verbos constituyen las ocurrencias de la muestra de trabajo (Véase Anexo A). Estos se tomaron conforme a la frecuencia de aparición en el total de la muestra; es decir, iniciando con el más frecuente (hacer con 78 apariciones) se filtró la lista de verbos hasta alcanzar el 50 % de los datos (avanzar con 4 apariciones).

Siguiendo la propuesta de análisis presentada en el artículo sobre *Cómo se representa lingüísticamente la pobreza en Colombia* (Quiroz et al., 2017, p. 83), los datos arrojados por la



frecuencia de los verbos, nos permitirían dar cuenta de su uso homogéneo con aquellos empleados en el Programa de las Naciones Unidas para el Desarrollo (PNUD). Lo que a su vez, permitiría establecer que lingüísticamente, los periódicos colombianos presentan una elección del verbo en la presentación de la pobreza, que coincide con este organismo multilateral; lo que genera una prelación por las verbos positivos sobre los negativos.

En la *web* del PNUD para Colombia, se presenta el informe al año 2015 donde se afirma que el país, ha superado los Objetivos de Desarrollo del Milenio (ODM) (desde el 2016 denominados Objetivos de Desarrollo Sostenible (ODS)). Con relación al ítem de pobreza “El país logró cumplir las metas de reducción de la pobreza fijadas para el 2015 con un año de anticipación”(Naciones Unidas, 2015), esto es consecuente con la constante que hemos identificado de elementos positivos o neutros frente a los negativos.

La figura 12 relaciona los 27 verbos más frecuentes según la productividad en las ocurrencias analizadas que dan cuenta del 50 % del total de los verbos. En los numerales del 65 al 70, se presentan ejemplos de los 6 verbos más productivos, los cuales dan cuenta del 20,3 % del total.

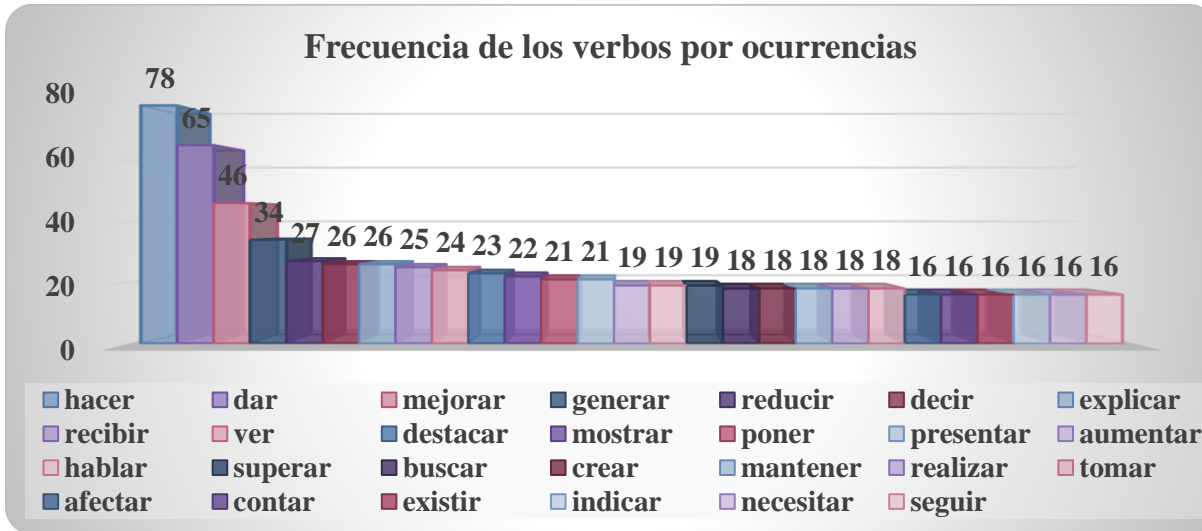


Figura 12: Frecuencia verbos por ocurrencias

65. **hacer** pronunciamiento

66. **dar** rienda suelta

67. **mejorar** calidad de vida

68. **generar** trabajo digno

69. **reducir** desocupación abierta

70. **decir** progreso

#### 6.1.4. FRECUENCIA SEGÚN EL TIPO DE VERBO

Producto de la tipificación realizada en este trabajo, con base a la propuesta de Lorente (2002), se da cuenta de una distribución de los verbos en tres categorías: fraseológicos (54 verbos); discursivos (22 verbos); y conectores (13 verbos). De igual modo, con relación a la distribución de las ocurrencias, de acuerdo al tipo de verbo, los fraseológicos son los más productivos, seguidos de los discursivos y en último lugar los conectores. En la tabla 21 se relaciona la cantidad de ocurrencias según el tipo de verbo. En los numerales 71 al 73 se presentan ejemplos según cada tipo de verbo.

Tabla 21: Frecuencia de las ocurrencias con relación al tipo de verbo

Tipo de verbo	Frecuencia	%
Fraseológico	827	60,81
Discursivo	324	23,82
Conector	209	15,37
Palabras o elementos distintos contados: 3		
Total	1.360	100%

71. *Fraseológicos*: aumentar, construir, enfrentar, fortalecer, implementar, salir, trabajar

72. *Conectores*: cambiar, cumplir, encontrar, llevar, dar, incluir, quedar, pedir, tomar

73. *Discursivos*: explicar, hablar, indicar, llamar, mostrar, presentar, resaltar, señalar, ver

A continuación, se presenta la información con relación al tipo de verbo, de acuerdo a los datos encontrada y teniendo como referencia la propuesta de Lorente (2002) descrita en el marco teórico (Capítulo 3 § 3.1.1.4.1.3.).

##### 6.1.4.1. VERBOS FRASEOLÓGICOS

En la figura 13 se presenta la frecuencia de aparición de los verbos fraseológicos. Se tomó el 75 % de los datos correspondiente a 32 verbos.

## Verbos fraseológicos

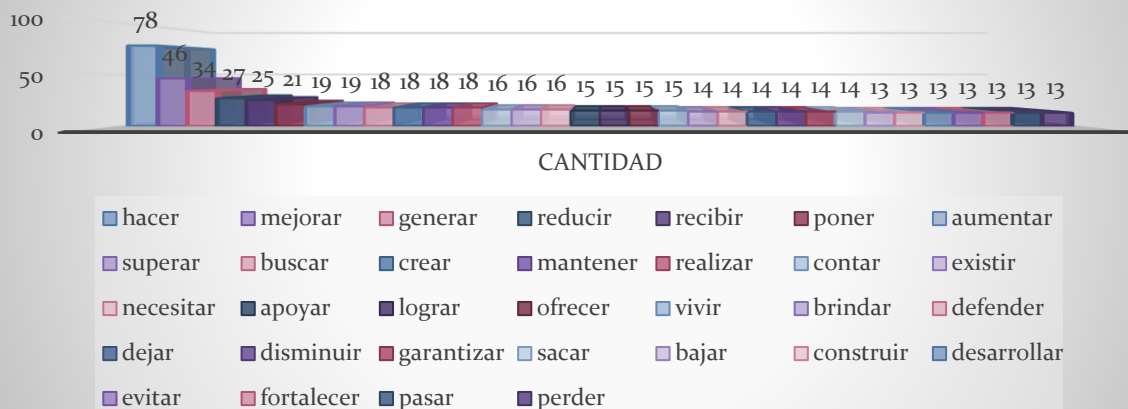


Figura 13: Verbos fraseológicos

Los verbos fraseológicos son los encargados de establecer el carácter técnico del discurso, los cuales según Lorente (2002), toman carácter de especialidad “cuando aparecen en contexto con una unidad terminológica” (Lorente, 2002, par.38). Unido a la clasificación de los tipos de verbos, también se realizó una clasificación de los objetos directos acorde a sus áreas temáticas (Capítulo 6 § 6.1.9.). De esta forma, se podría establecer el carácter técnico de los textos analizados. En los numerales 74 al 78 se listan algunos ejemplos con este tipo de verbos.

74. **hacer** redistribución de la riqueza

75. **mejorar** calidad de la educación

76. **generar** empleo formal

77. **reducir** índice de pobreza extrema

78. **recibir** alimentación

### 6.1.4.2. VERBOS CONECTORES

Como se ha venido explicando, según Lorente (2002), los verbos conectores ayudan en la construcción del conocimiento especializado, más no son poseedores del mismo. En esta línea, a pesar de ser constructores del aspecto técnico del texto (en el diseño de las herramientas para el Procesamiento de Lenguaje Natural (PLN)), debe anotarse su uso exclusivo como puente en la

construcción del conocimiento específico. En la figura 14 se presentan los verbos conectores junto a ejemplos con este tipo de verbo en 79 a 83.

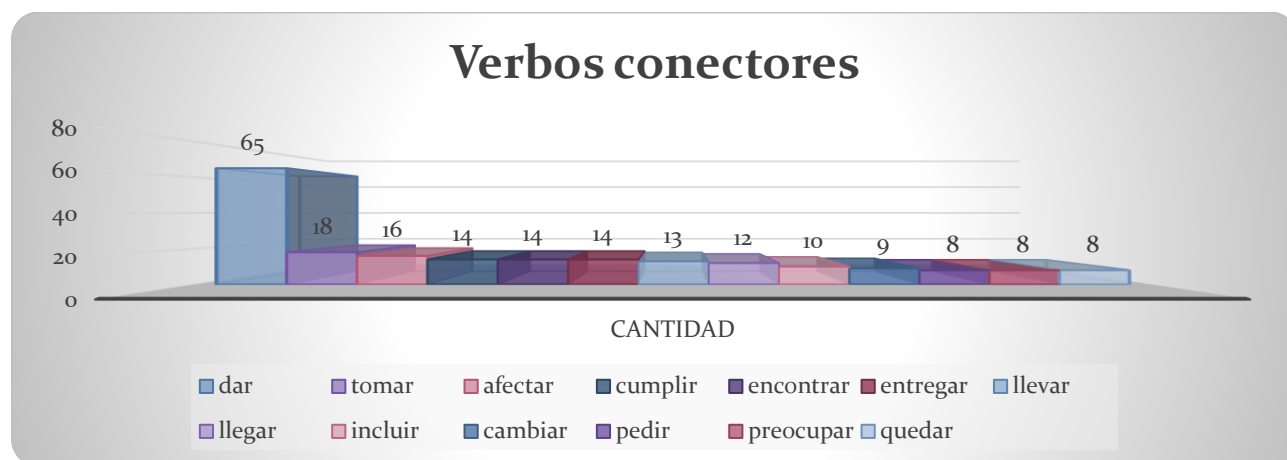


Figura 14: Verbos conectores

79. **dar** incentivo

80. **tomar** IMP

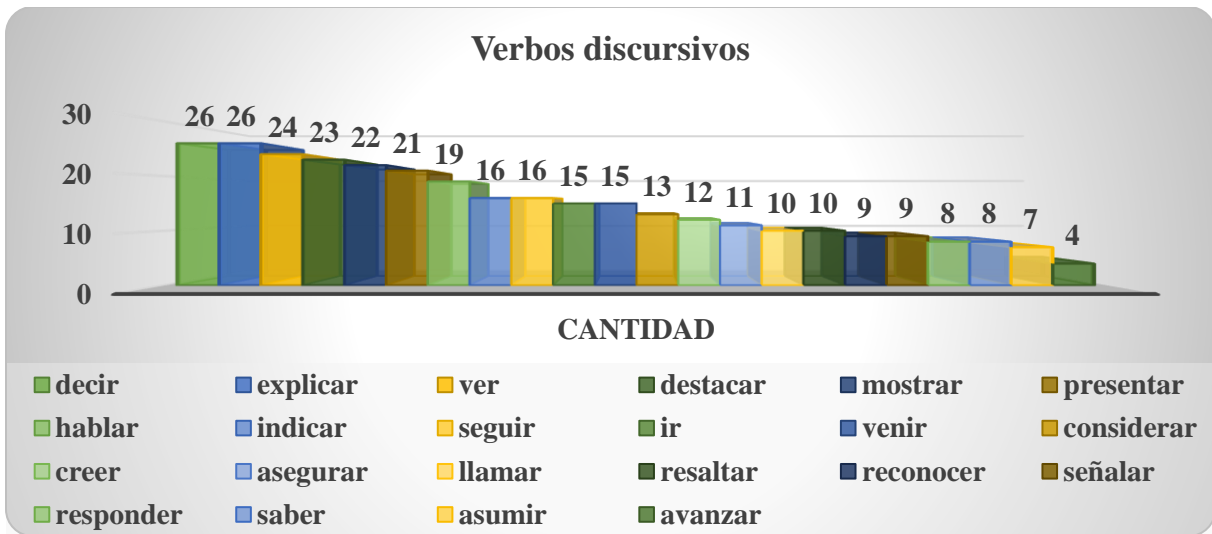
81. **afectar** indicador nacional

82. **cumplir** objetivo

83. **encontrar** crecimiento

#### 6.1.4.3. VERBOS DISCURSIVOS

En cuanto a los verbos discursivos, su principal función es de tipo estilístico, debido a que reúnen las estrategias en la construcción del discurso desde elementos de coherencia y cohesión, al igual que introducen las ideas expresadas por los verbos fraseológicos y conectores. En la figura 15 se presentan los porcentajes de distribución de los verbos discursivos. De igual forma, se presentan ejemplos en los numerales 84 al 88 con ocurrencias con este tipo de verbos.



*Figura 15: Verbos discursivos*

- 84. **decir** informe de prensa
- 85. **explicar** objetivo del programa
- 86. **hablar** prensa
- 87. **indicar** punto de referencia
- 88. **mostrar** oportunidad de crecimiento

Teniendo de presente, lo descrito sobre los verbos fraseológicos y conectores (76,18 % del total de los verbos), como constructores del conocimiento específico (en la categorización de Lorente (2002)). Se podría concluir que el tema de la pobreza en los periódicos analizados, es abordado con un énfasis técnico.

### 6.1.5. CARGAS SEGÚN EL TIPO DEL VERBO

A la hora de recuperar la información correspondiente a la relación entre el tipo de verbo con la carga asignada, se obtuvieron datos importantes en la construcción de una tipología del neutro; la cual, por motivos metodológicos ha sido constantemente excluida de las herramientas para AS (Liu, 2015). Lo anterior, ha dejado un vacío conceptual con relación a lo que se puede denominar como neutro. La tabla 22 reúne los datos con relación a la frecuencia de las ocurrencias según la tipología de los verbos conforme a la carga asignada a estos.

Tabla 22: Frecuencia cargas ocurrencias según tipo del verbo

Tipo de verbo	Fraseológico		Discursivo		Conector	
	Frecuencia	%	Frecuencia	%	Frecuencia	%
Positivo	411	49,7	0	0	89	42,58
Negativo	140	16,93	0	0	24	11,48
Neutro	276	33,37	324	100	96	45,93
Palabras o elementos distintos contados:	3		3		3	
Total:	827		324		209	

Con respecto a la polaridad neutra, se encontraron dos tendencias en la literatura del área, las cuales se han analizado a la luz de nuestros datos, con miras a establecer una definición de neutro producto de este trabajo de investigación.

En primera instancia, en *Lexicon-Based Methods for Sentiment Analysis* Taboada, Brooke, Tofiloski, Voll, y Stede (2011), afirman que las “*words with SO = 0 were sometimes interpreted as positive, but almost never interpreted as negative*”, es decir: “*neutral description might be taken as being vaguely positive, but it would not be mistaken for negative expression*” (p.290). En segunda instancia, en “*The dynamics of seller reputation: Evidence from eBay*” Cabral y Hortaçsu (2004) afirman que lo neutro está más cerca de lo negativo:

*Our subjective impression, after browsing through eBay community chatboards where users discuss issues regarding the feedback system, is that the information contained in a neutral rating is perceived by users to be much closer to negative feedback than positive* (p.6).

Adicionalmente, en Tokuhiya, Inui, & Matsumoto (2008), aparte de plantear el desafío y la dificultad metodológica de una clasificación triclase (positivo, negativo y neutro), establecen que las “*positive sentences tend to be classified as neutral*” (p.886).

La tendencia en este proyecto de investigación está en coherencia con lo planteado por Taboada *et al.*(2011). Las ocurrencias marcadas como neutras están más cerca de ser consideradas positivas que negativas, lo cual se va corroborando a medida en que se avanza en la investigación. A continuación, se describe la distribución de los verbos según: la carga asignada y su clasificación. Para este análisis se tomó la cantidad de verbos analizados (89) más no la frecuencia de ocurrencias con ellos construidas.

### 6.1.5.1. POSITIVO

Acorde a los datos presentados en la tabla 23, sobre la distribución de los verbos positivos según el tipo, se ha concluido que los verbos fraseológicos son los discriminantes en la polaridad.

Tabla 23: Verbos fraseológicos según carga de polaridad

<b>Verbos positivos</b>		
<b>Tipo</b>	<b>Frecuencia</b>	<b>%</b>
Fraseológico	28	90,32
Conector	3	9,68
Palabras o elementos distintos contados: 2		
Total:	31	100

### 6.1.5.2. NEGATIVO

La tabla 24 refiere la distribución de los verbos con carga negativa, en concordancia con lo expresado sobre el carácter discriminante de los verbos fraseológicos. La distribución de los verbos negativos da cuenta de un comportamiento igual; es decir, para estos verbos, la discriminación de la polaridad recae sobre los clasificados como fraseológicos.

Tabla 24: Verbos conectores según carga de polaridad

<b>Verbos negativos</b>		
<b>Tipo</b>	<b>Frecuencia</b>	<b>%</b>
Fraseológico	10	83,33
Conector	2	16,67
Palabras o elementos distintos contados: 2		
Total: 12		

### 6.1.5.3. NEUTRO

En la tabla 25 se relaciona la información sobre los verbos marcados como neutros según el tipo. En este punto, los resultados han permitido identificar dos elementos determinantes, con miras a ofrecer una posible definición de lo neutro como categoría de polaridad lo que constituiría un aporte al área del AS.

Tabla 25: Verbos conectores según carga de polaridad

<b>Verbos neutros</b>		
<b>Tipo</b>	<b>Frecuencia</b>	<b>%</b>
Discursivo	22	47,83
Fraseológico	16	34,78
Conector	8	17,39
Palabras o elementos distintos contados: 3		
<b>Total:</b>	46	100

Los verbos marcados como discursivos fueron etiquetados con carga neutra, lo que en primera instancia significa que el neutro, como variable de polaridad, aparece en el uso del lenguaje con fines comunicativos. Esto implica, que el uso de este tipo de verbos está ligado a la construcción del discurso desde estrategias textuales como la coherencia y la cohesión; lo que se aleja de la idea de comunicar la opinión del escritor; siendo este aspecto lo que se busca identificar a la hora de llevar a cabo la clasificación de textos mediante la aplicación de las herramientas para el AS.

De este modo, con relación a Lorente (2002) sobre este tipo de verbos; la representación léxica de estos en las herramientas de procesamiento del lenguaje debería incluir una anotación de su uso meramente discursivo.

En segunda instancia, con relación a las características ya definidas de los verbos fraseológicos y conectores, como constructores del discurso especializado y no como portadores de este. El neutro se define con relación a los elementos que no son temáticos en el sentido que se ha descrito; es decir, las unidades que no son terminológicas. En suma, acorde a los datos analizados y descritos, se presenta una posible definición de lo neutro como: *los elementos textuales no poseedores de carácter temático ni especializado, cuya función es discursiva; toda vez, que no aportan a la polaridad del texto, sino que ayudan en la construcción del discurso, desde las estrategias de coherencia y cohesión.*

#### **6.1.6. FRECUENCIA SEGÚN EL SENTIDO DEL VERBO**

Conforme a la categorización utilizada en *WordNet* que, a su vez, está basada en la propuesta de Levin (1993), hemos identificado un total de 14 sentidos verbales en los datos. La figura 16 ilustra la frecuencia de aparición de acuerdo con los sentidos en las ocurrencias analizadas.



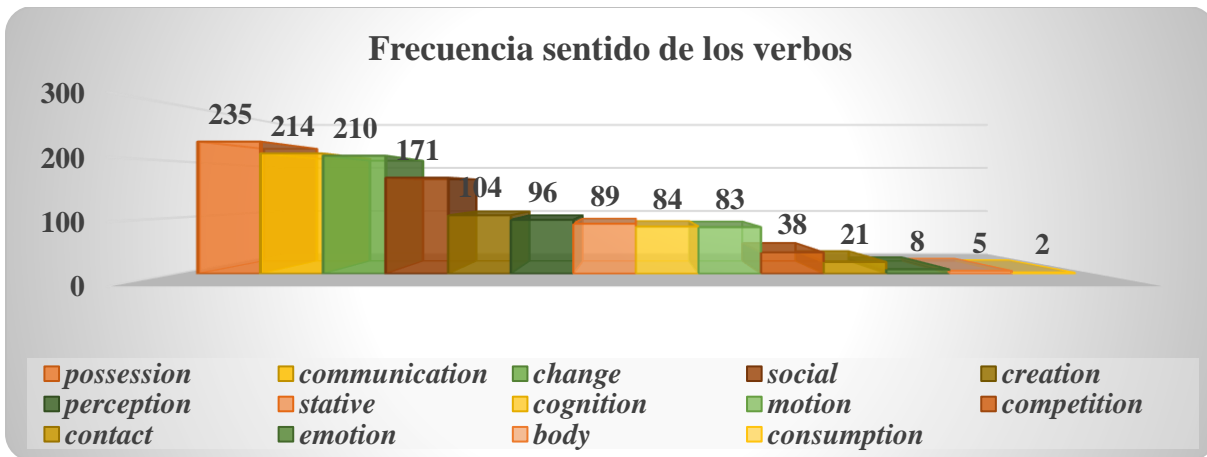


Figura 16: Frecuencia sentido de los verbos

Se ha conservado el nombre de las categorías en inglés con la finalidad de evitar ambigüedades o imprecisiones en el manejo de las etiquetas y permitir de igual forma, la comparación de los resultados aquí ofrecidos con otros trabajos en el campo.

¿Qué información aporta el sentido de los verbos? El sentido de los verbos define la forma como es tematizado el problema, es decir, la manera como se aborda el fenómeno desde las áreas o elementos temáticos implicados y con relación a los aspectos que confluyen en el desarrollo de un esquema alrededor de la descripción y evolución del tema en cuestión. Unido a esto, el sentido de verbo más frecuente corresponde a los de *possession* con un porcentaje de 17,28 %; mientras los primeros 6 sentidos, de los 14 hallados en los datos, dan cuenta del 75,74 % de los datos.

Siguiendo la línea de los Objetivos del Milenio (Naciones Unidas, 2000) con relación a la pobreza, estos apuntan a su erradicación mediante políticas que permitan, en gran medida, encontrar un empleo digno y estable; y a su vez, la cobertura de las necesidades elementales de bienes y servicios vitales. Esto favorece salir del estado de pobreza y, por lo tanto, la erradicación de la misma. Podría establecerse que la elección de los verbos en la representación del fenómeno en la prensa colombiana está en directa relación con la consecución de estos objetivos, lo que permite entrever que, al menos lingüísticamente, la prensa colombiana sigue la misma línea del PNUD y de allí que estos se conviertan en herramientas para difundir ideologías políticas, económicas o culturales (Quiroz et al., 2017, p. 84), confirmando la manipulación de la información en los medios ya evidenciada desde (Chareau, 1997).

Así mismo, en el comunicado de prensa al respecto de la pobreza en Colombia, del Departamento Administrativo Nacional de Estadística (DANE) (2012) (entidad colombiana responsable de la planeación, levantamiento, procesamiento, análisis y difusión de las estadísticas oficiales de Colombia), se presenta la dos formas como desde el año 2012 se viene midiendo la pobreza en el país: la primera, según el cálculo de los ingresos monetarios (pobreza monetaria) y, la segunda, a partir de 5 dimensiones y 15 criterios (Índice de Pobreza Multidimensional, IMP) a partir de la propuesta de Alkire y Foster en “*Counting and Multidimensional Poverty Measurement*” (2008). Estas dimensiones y criterios son descritos en detalle en el comunicado de prensa arriba mencionado. Para los análisis de nuestra investigación, tomaremos las dimensiones y criterios del IMP.

La figura 17 ilustra las 5 dimensiones en las cuales se basa el IMP, al igual que los 15 criterios allí analizados. En Colombia, una persona se considera pobre cuando el porcentaje de privación en los ítems analizados es del 33 %.

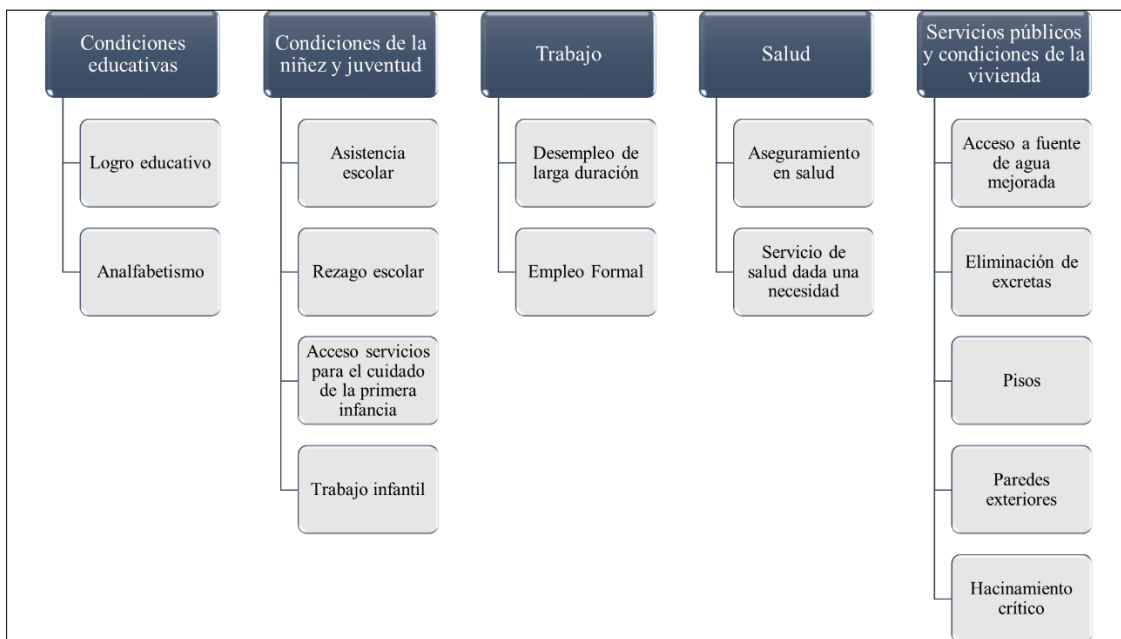


Figura 17: Dimensiones del IMP en Colombia

Los tipos de verbos que hemos categorizado en nuestros datos, dan cuenta de estas cinco dimensiones al indicar cómo la dinámica en Colombia, entre el 2010 y 2014 periodo en el que se ubican los artículos de nuestro corpus, ha superado continuamente los índices según los cuales una persona es considerada en condición de pobreza, como hemos constatado en los ejemplos

recuperados de nuestros datos y listados junto a una descripción de cada sentido verbal, con frecuencia superior al 5 %. Por último, tomamos como referencia el trabajo de Levin (1993) para definir estos 9 sentidos de verbos (*possession, communication, change, social, creation, perception, stative, cognition* y *motion*) que superan el punto de corte establecido.

### 6.1.6.1. POSSESSION

En la clasificación ofrecida por Levin (1993), los verbos de *possession* son tipificados bajo la categoría de *verbs of change of possession* y *verbs of possessional deprivation*. De acuerdo con esta autora, los verbos agrupados en este sentido expresan por un lado la pérdida de posesión puesto se entienden como “*the removal of something from someone’s possession*” (p.129) o “*depriving someone/something of an inalienable possession*” (pp.129-130) y, por otro lado, también expresan el *change of possession*, en cuanto dan como resultado, la transferencia de la posesión desde varios puntos de vista. De allí que este sentido de verbo sea transversal a otros, a saber, *debone verbs, send verbs, slide verbs, bring and take verbs*, entre otros.

Con relación a nuestros datos, hemos encontrado 235 ocurrencias con verbos de sentido *possession*, equivalentes al 17, 28 % construidas con un total de 16 verbos. La figura 18 relaciona los verbos cuya frecuencia es superior al 5 %. La frecuencia de los verbos por debajo de este porcentaje es inferior a 10 ocurrencias.

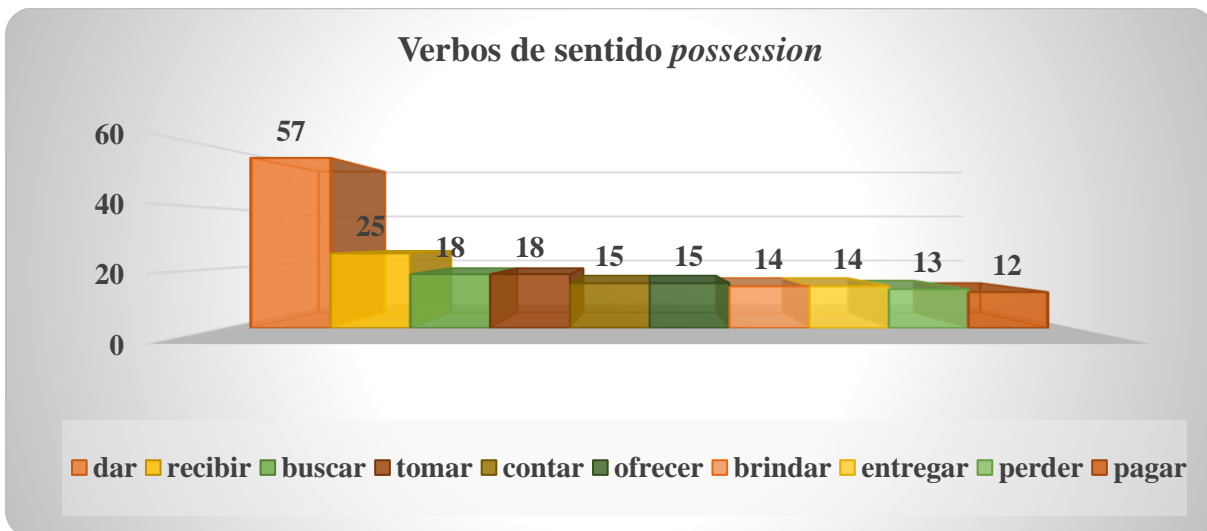


Figura 18: Frecuencia verbos de sentido *possession*

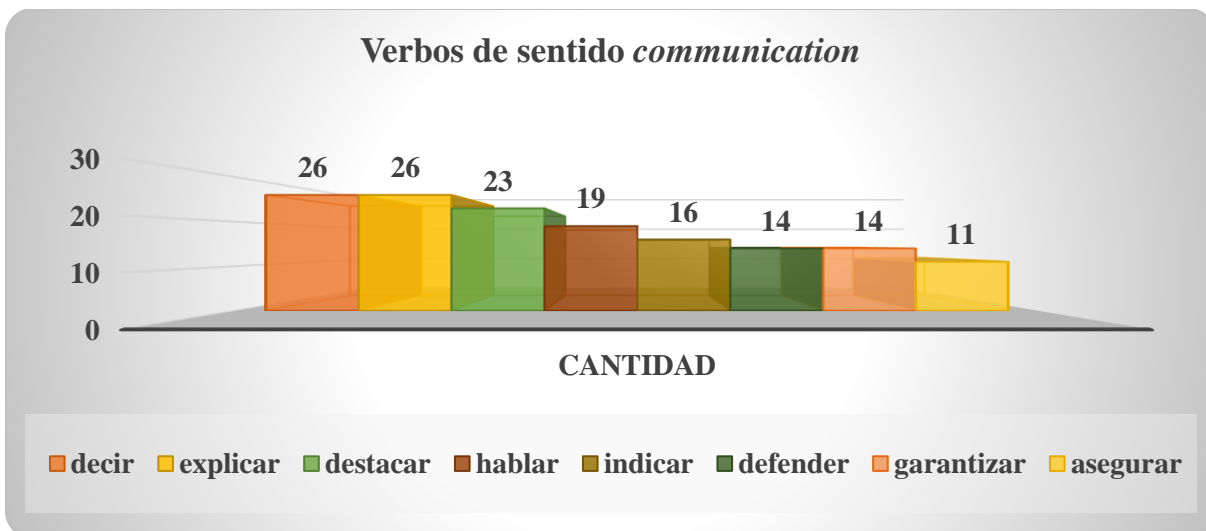
Al ser verbos de *possession*, hemos constatado que con ellos se describe precisamente la posesión con relación a los criterios del IMP, es decir, las construcciones con este tipo de verbo dan cuenta del cómo se cumple con estos criterios, como se evidencia en los ejemplos listados en 89 a 98.

- 89. **dar** educación
- 90. **recibir** pensión
- 91. **buscar** trabajo
- 92. **tomar** predio
- 93. **contar** agua potable
- 94. **ofrecer** trabajo
- 95. **brindar** asistencia técnica
- 96. **entregar** mercado
- 97. no **perder subsidio**
- 98. **pagar** deuda

#### **6.1.6.2. COMMUNICATION**

Bajo este sentido, se han agrupado los verbos que corresponden a la “*communication and transfer of ideas*” (Levin, 1993, p.202). De acuerdo a la categorización ofrecida por la autora, estos verbos se agrupan en 8 subcategorías: *verbs of transfer*, *verbs of manner of speaking*, *verbs of instrument of communication*, *talk verbs*, *chitchat verbs*, *say verbs*, *complain verbs* y *advise verbs*; las cuales no son descritas sistemáticamente por la autora puesto que, según afirma, sobrepasan los límites de su investigación.

En nuestros datos, hemos recuperado 214 ocurrencias con verbos de sentido *communication*, correspondientes al 15,74 %, las cuales se construyen con un total de 17 verbos. La figura 19 relaciona los verbos cuya frecuencia es superior al 5 %. La frecuencia de los verbos por debajo de este porcentaje es inferior a 10.



*Figura 19: Frecuencia verbos de sentido *communication**

La mayor parte de las ocurrencias aquí listadas, corresponden a las estrategias discursivas del escritor donde se busca introducir los temas a abordar o simplemente conectar las ideas al interior del texto. Los verbos en este grupo son en su mayoría de tipo discursivos. En los numerales 99 a 106 se muestran ejemplos con verbos de sentido *communication*.

- 99. **decir** comunicado de prensa
- 100. **explicar** nuevo acuerdo
- 101. **destacar** avance logrado
- 102. **hablar** política agraria
- 103. **indicar** ingreso real
- 104. **defender** proceso de paz
- 105. **garantizar** objetivo constitucional
- 106. **asegurar** anhelo de paz

### 6.1.6.3. *CHANGE*

En cuanto a los verbos de sentido *change*, Levin los presenta con relación a cambios en distintos aspectos como cambio de significado o de estado y, a su vez, verbos que producen estos cambios. Este sentido de verbo es transversal a varias categorías y son clasificados, según la autora,

en verbos de *change of bodily state* o *change of state* y están presentes en las demás categorías en las cuales son transversales, como en los verbos de *possession* (listados previamente).

Con un total de 17 verbos se construyen las 210 ocurrencias para el sentido *change*, correspondientes al 15,44 % de los datos. La figura 20 relaciona los verbos cuya frecuencia es superior al 5 %. La frecuencia de los verbos por debajo de este porcentaje es inferior a 10.

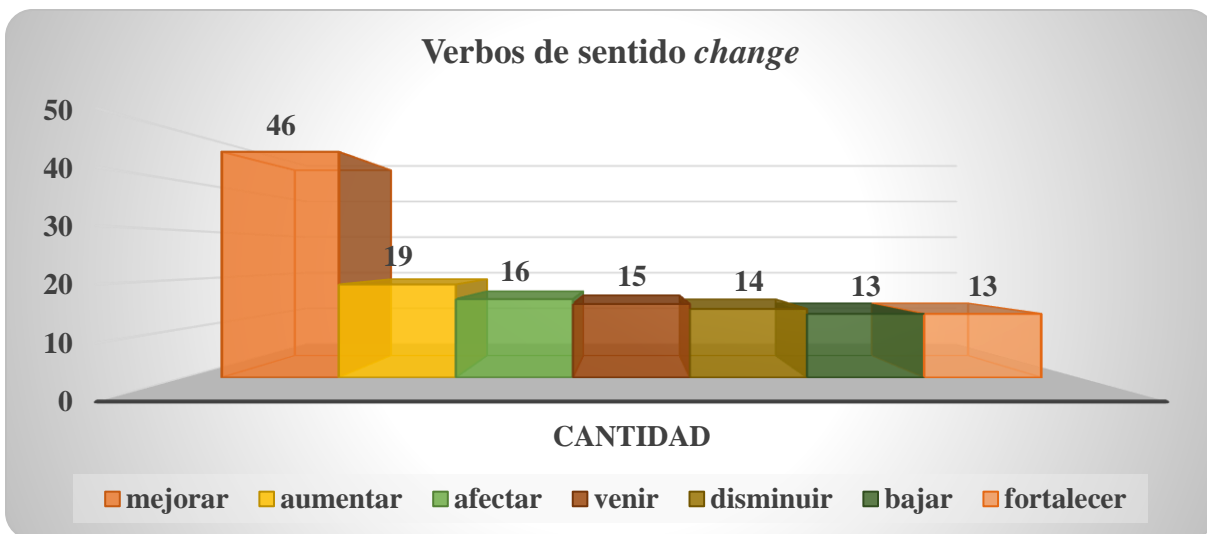


Figura 20: Frecuencia verbos de sentido *change*

Con este tipo de verbos, hemos hallado claramente cómo es representada la dinámica con relación a la pobreza. Los verbos más frecuentes en esta categoría, mayormente fraseológicos (13 de 17), en unión con su objeto directo, expresan una mejoría constante con relación a los criterios del IMP ya mencionados. En los numerales 107 a 113 se muestran ejemplos de ocurrencias con verbos de sentido *change* que dan cuenta de la tendencia que hemos descrito reiteradamente.

- 107. **mejorar** el ranking mundial
- 108. **aumentar** empleo
- 109. **afectar** competitividad
- 110. **venir** mercado externo
- 111. **disminuir** pobreza total
- 112. **bajar** precio medicamento

113. **fortalecer** seguridad alimentaria

#### 6.1.6.4. *SOCIAL*

Para Levin, estos verbos corresponden principalmente a lo que ella denomina *social interaction*, en el sentido de las actividades que involucran más de un participante. Verbos relacionados con *fighting* o *verbal interactions* son agrupados bajo esta categoría (p.36). Otra característica importante sobre este sentido de verbos, expresa la autora, es que cuando se refieren a un sujeto individual, requieren un objeto directo o un sintagma preposicional a partir de la proposición *con*.

En nuestros datos hemos recuperado 171 ocurrencias con verbos de sentido *social*, correspondientes al 12,57 % de los datos, construidas con un total de 10 verbos. La figura 21 presenta los verbos cuya frecuencia es superior al 5 %. La frecuencia de los verbos por debajo de este porcentaje es inferior a 8.

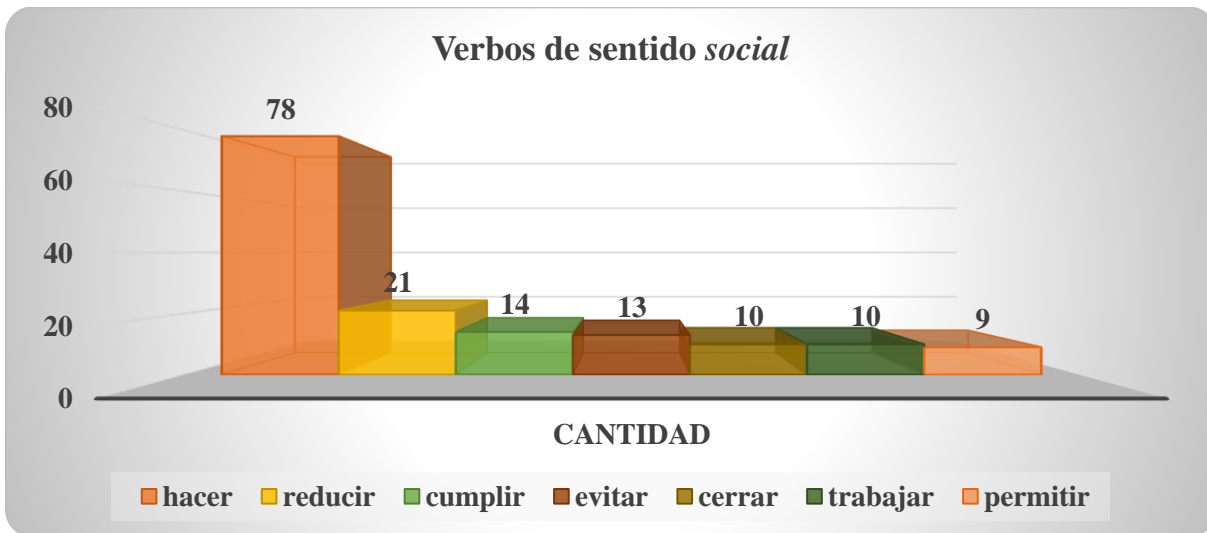


Figura 21: Frecuencia verbos de sentido *social*

Con los verbos de sentido *social* se hace referencia al fenómeno, pero desde la perspectiva del sujeto, esto es, se analiza el fenómeno, pero en consonancia con la persona o grupo de personas aludidas. En los numerales 114 a 120 se muestran ejemplos de ocurrencias con verbos de sentido *social* para ilustrar su uso.

114. **hacer** parte del parte del proyecto de VIP

- 115. **reducir** hambre
- 116. **cumplir** compromiso
- 117. **evitar** quiebra
- 118. **cerrar** brecha de inequidad
- 119. **trabajar** inclusión social
- 120. **permitir** distribución equitativa

**6.1.6.5. CREATION**

Para Levin, estos verbos son agrupados bajo la denominación de *creation and transformation*. La autora precisa que la mayoría de los verbos aquí agrupados son transitivos, en cuanto retoman como argumento un agente que crea o transforma la entidad (p.172); es decir, los verbos de *creation* son aquellos que indican, precisamente, creación o transformación.

El total de 104 ocurrencias de verbos de sentido *creation* se construyen con 7 verbos. La figura 22 relaciona los verbos cuya frecuencia es superior al 5 %. Para esta categoría solo se cuenta con un verbo (dar) por debajo de este porcentaje, con un total de 2 ocurrencias (1,92 %).

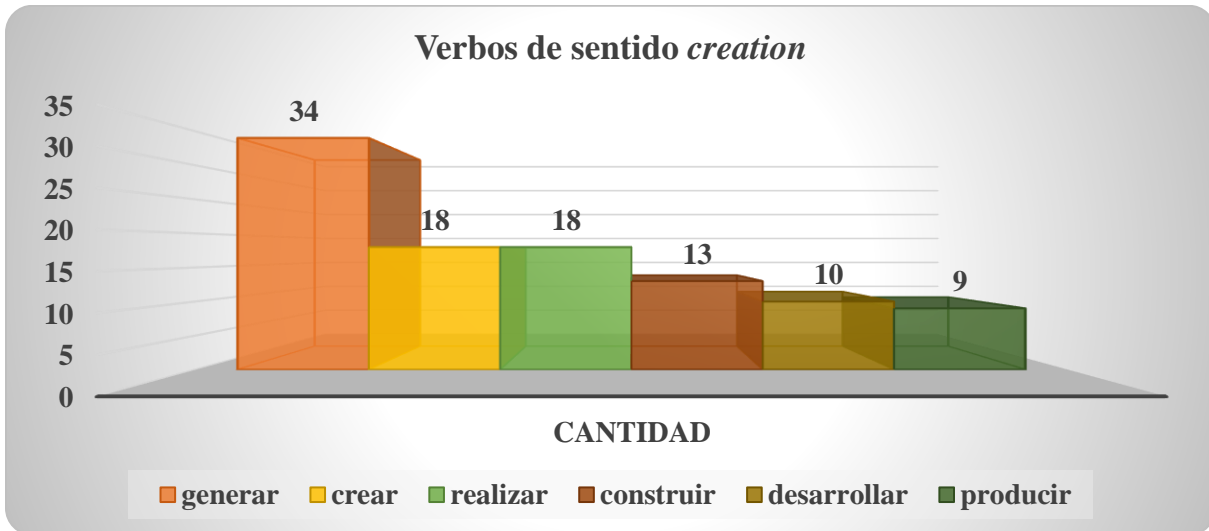


Figura 22: Frecuencia verbos de sentido *creation*

A la par con los verbos de *change*, los verbos agrupados en este sentido aluden a la mejoría constante con relación a los criterios del IMP; es decir, con estos verbos se recuperan los elementos



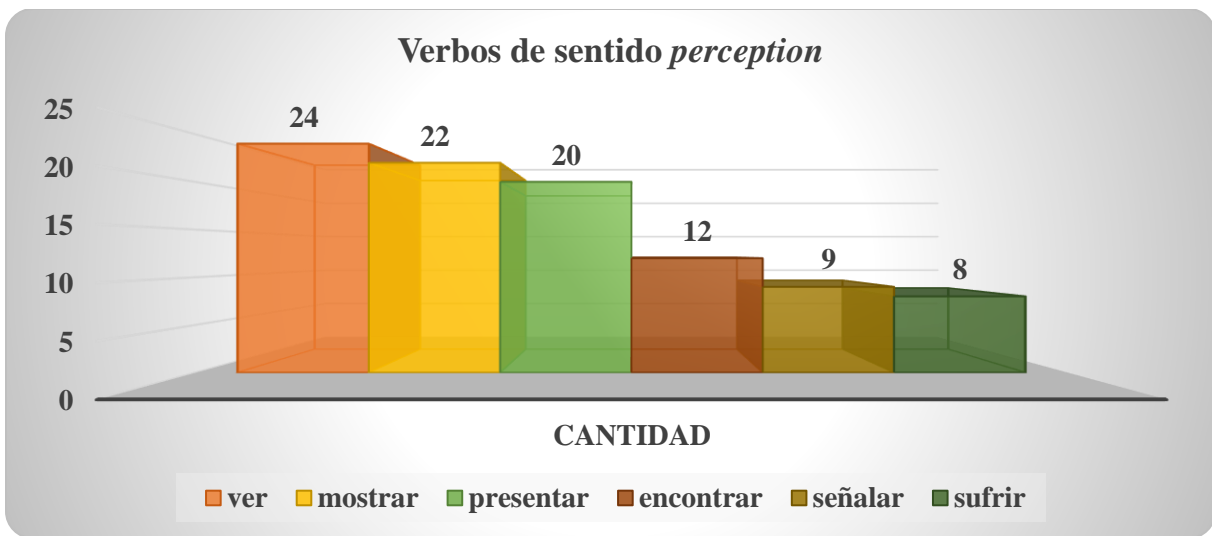
referidos a la implementación de estrategias con miras a superar los índices con relación a los criterios establecidos en el IMP. En los numerales 121 a 126 se muestran ejemplos de verbos con sentido *creation*.

- 121. **generar** empleo formal
- 122. **crear** política pública masiva
- 123. **realizar** curso de cocina básica
- 124. **construir** vivienda
- 125. **desarrollar** trabajo calificado
- 126. **producir** resultado positivo

#### 6.1.6.6. **PERCEPTION**

En Levin, los verbos de *perception* dan cuenta de la relación entre el que percibe (sujeto) y lo que es percibido (objeto), en el caso de los verbos transitivos; en cuanto estos “*describe the actual perception of some entity*” (p.186) y son agrupados en los *see* y *sight* verbs. Para el caso de los verbos intransitivos, estos toman un complemento preposicional y son agrupados en las categorías de *peer* y *stimulus subject perception verbs*.

Las ocurrencias con verbos de sentido *perception* equivalen al 7,06 % de los datos y se construyen con un total de 7 verbos. La figura 23 relaciona los verbos cuya frecuencia es superior al 5 %. Al igual que los verbos de sentido *creation*, esta categoría solo cuenta con un verbo (sacar) por debajo de este porcentaje con un total de 1 ocurrencia (1,04 %).



*Figura 23: Frecuencia verbos de sentido *perception**

Como se ve en los numerales 127 a 132, los ejemplos de ocurrencias con verbos de sentido *perception* indican, al igual que los verbos de *communication*, elementos de cómo se percibe el fenómeno; es decir, indican elementos con relación a la evolución de la pobreza teniendo siempre como referente, los criterios del IMP. De nuevo, las ideas planteadas a partir de estos criterios, indican en mayor proporción, mejoría con relación al referente mencionado.

- 127. **ver** mejoría
- 128. **mostrar** igualdad
- 129. **presentar** nivel bajo de pobreza
- 130. **encontrar** crecimiento
- 131. **señalar** crecimiento económico
- 132. no **sufrir** enfermedad

#### **6.1.6.7. *STATIVE***

Los verbos de sentido *stative*, en la clasificación presentada por Levin, están enmarcados en *the resultative construction*, la autora indica que este tipo de verbos, debido a sus restricciones semánticas, han sido excluidos del análisis. Sin embargo, al hablar de *resultative phrase*, la autora hace claridad que “A *resultative phrase* is an *XP* which describes the state achieved by the referent

of the noun phrase it is predicated of as a result of the action named by the verb” (p.101), donde XP hace referencia a las *X’s way construction* abordadas, por Levin, en la *section 7.4* de su trabajo.

Asimismo, con los verbos de estado, el sustantivo que los acompaña, ante la ausencia de estos, hace las veces de verbo; es decir, asumen la carga semántica de este, como se evidencia en los ejemplos listados en los numerales 133 a 135.

133. **mantener** desarrollo agrícola

134. **existir** empleo

135. **quedar** remodelación

Las 89 ocurrencias de los verbos de sentido *stative* corresponden al 6,54 % de los datos y se construyen con un total de 9 verbos. La figura 24 relaciona los verbos cuya frecuencia es superior al 5 %. La frecuencia de los verbos por debajo de este porcentaje, es inferior a 3.

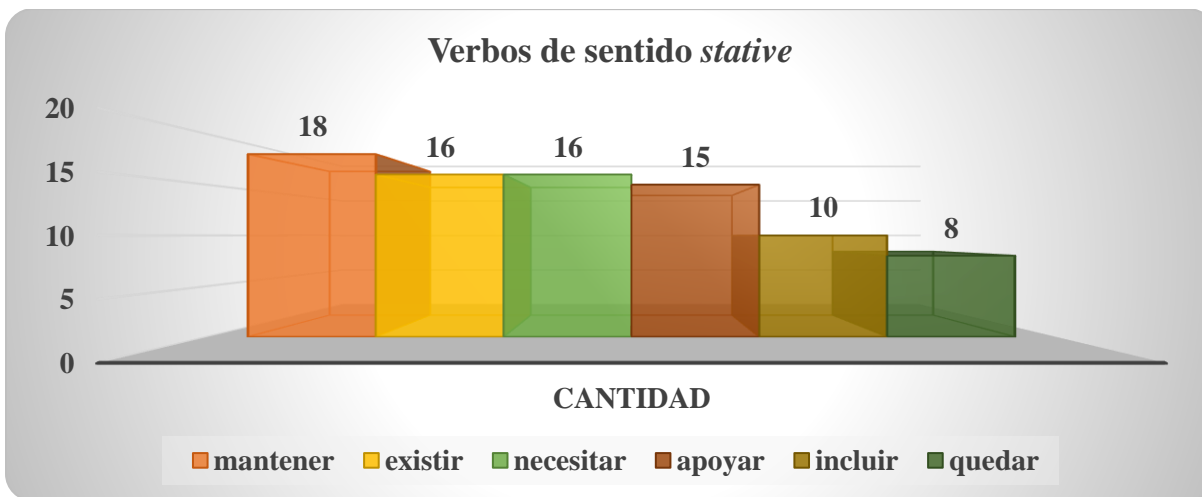
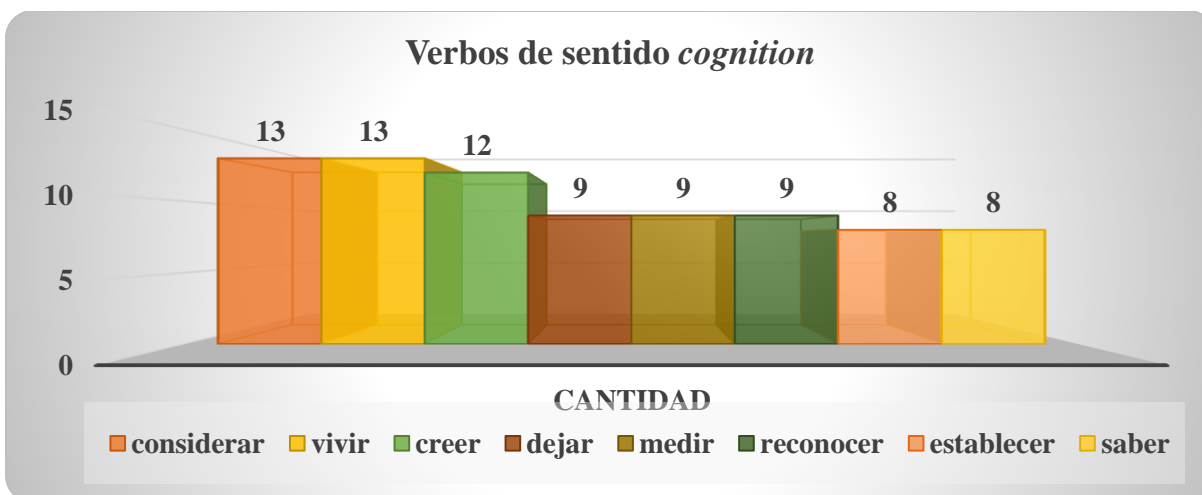


Figura 24: Frecuencia verbos de sentido *stative*

#### 6.1.6.8. *COGNITION*

Las 84 ocurrencias de verbos de sentido *cognition* corresponden al 6,18 % de los datos y se construyen con un total de 9 verbos. La figura 25 relaciona los verbos cuya frecuencia es superior al 5 %. De nuevo, para esta categoría, solo se cuenta con un verbo (dar) por debajo de este porcentaje, con un total de 3 ocurrencias (3,57 %).



*Figura 25: Frecuencia verbos de sentido *cognition**

Con los verbos de *cognition* se expresan principalmente ideas con relación al conocimiento del fenómeno; es decir, se da cuenta de acciones con relación a la identificación de los elementos relacionados con la pobreza. En los numerales 136 a 143 se muestran ejemplos de ocurrencias con verbos de sentido *cognition*.

- 136. **considerar** ganancia
- 137. **vivir** cobertura
- 138. **creer** política regional
- 139. **dejar** pobreza extrema
- 140. **medir** índice de concentración de la tierra
- 141. **reconocer** avance
- 142. **establecer** sector agropecuario
- 143. **saber** calidad de la educación

Es importante recordar que, para un mismo verbo, pueden existir distintos sentidos de acuerdo a la información sobre la categorización descrita en la documentación de *WordNet*. En nuestros datos, este verbo ha sido etiquetado con los sentidos de *cognition*, *consumption*, *creation*, *motion*, *possession*; como se muestra en la tabla 26. En los numerales 144 a 148 se listan ocurrencias con el verbo *dar* según lo distintos sentidos asignados.

Tabla 26: Frecuencia sentidos verbo dar

<b>Sentidos verbo dar</b>		
<b>Sentido</b>	<b>Frecuencia</b>	<b>%</b>
<i>possession</i>	57	87,69
<i>cognition</i>	3	4,62
<i>consumption</i>	2	3,08
<i>creation</i>	2	3,08
<i>motion</i>	1	1,54
Palabras o elementos distintos contados: 5		
Total	65	100

- 144. **dar** incentivo (*possession*)
- 145. **dar** oportunidad (*cognition*)
- 146. **dar** café (*consumption*)
- 147. **dar** luz (*creation*)
- 148. no se **dar** marcha (*motion*)

#### 6.1.6.9. **MOTION**

Los verbos de sentido *motion*, en Levin (1993), expresan movimiento de entidades animadas o inanimadas, indicando la dirección del movimiento o la manera como este se da. Con los verbos recuperados de nuestros datos, agrupados en este sentido verbal, se hace referencia a acciones llevadas a cabo para cumplir con los criterios establecidos en el IMP, como se ejemplifica en los numerales 149 a 154.

Los verbos de sentido *motion* presentan un total de 83 ocurrencias (6,1 % de los datos) construidas con un total de 9 verbos. La figura 26 relaciona los verbos cuya frecuencia es superior al 5 %. Los verbos por debajo de este porcentaje tienen una frecuencia inferior a 2.

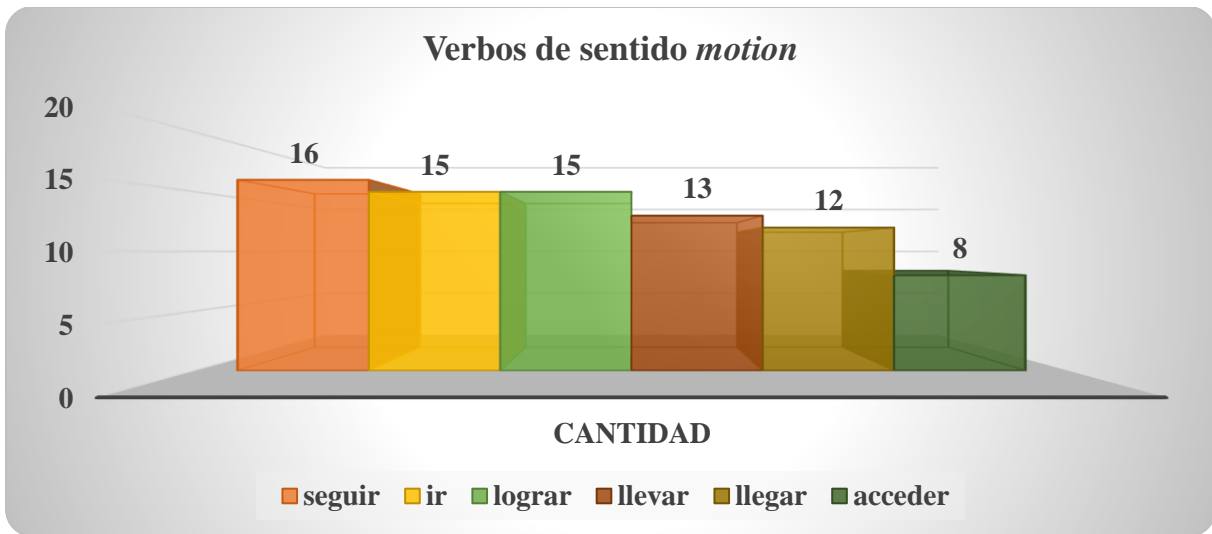


Figura 26: Frecuencia verbos de sentido *motion*

- 149. **seguir** cobertura
- 150. **ir** manejo del dinero
- 151. **lograr** transformación profunda
- 152. **llevar** gas
- 153. **llegar** paz total
- 154. **acceder** seguridad social

### 6.1.7. CARGAS DE LOS VERBOS

A cada uno de los verbos se le asignó una carga, conforme a las tres posibles polaridades en nuestro trabajo (positiva, negativa o neutra). La información de este etiquetaje permitiría identificar la elección léxica de los verbos para construir las referencias sobre el tema; es decir, por medio de la carga del verbo, hemos hallado que los etiquetados como positivos y neutros constituyen el 87,94 % de los datos mientras el 12,06 % restante pertenece a los verbos negativos. Asimismo, con este marcaje se establecieron las posibles combinaciones que ayudarían a identificar la forma como interactúan los elementos del patrón en los textos analizados. Estas combinaciones se explican en detalle más adelante (Capítulo 6 § 6.1.1.1.).

A su vez, al ser la pobreza un tema negativo, es a partir de la carga de los verbos que se podría evidenciar con mayor claridad el uso de construcciones positivas en la prensa colombiana, en la estructuración y posterior presentación de la pobreza. Además, al mirar en detalle las ocurrencias construidas con verbos negativos, hemos identificado que, aunque la diferencia es mínima, las construcciones positivas son más frecuentes y al sumarlas con las neutras, constituyen el 55,49 % de las ocurrencias con verbo negativo.

La figura 27 da cuenta de la frecuencia de estas con respecto a la carga del verbo que las construye.

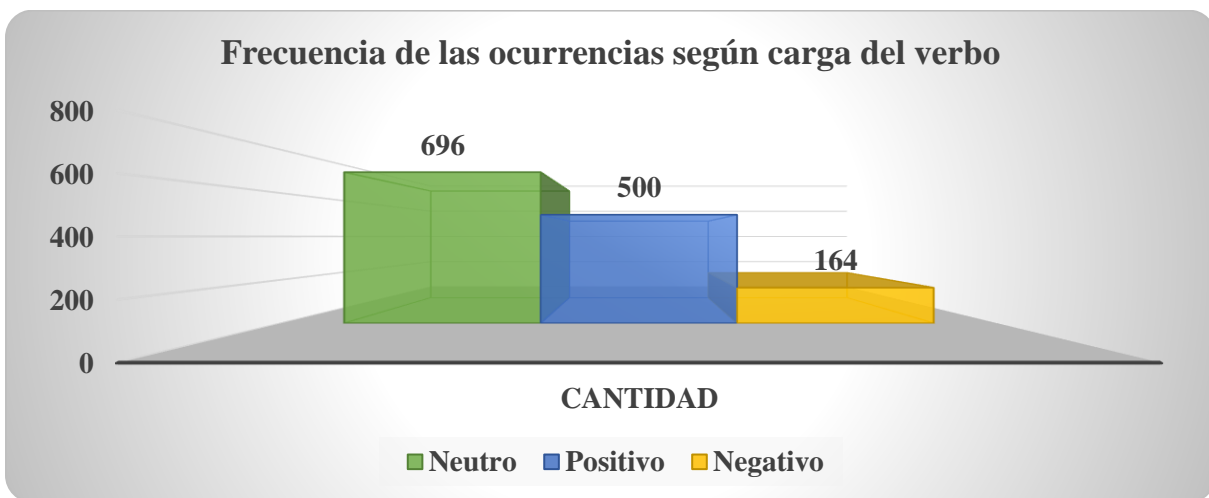
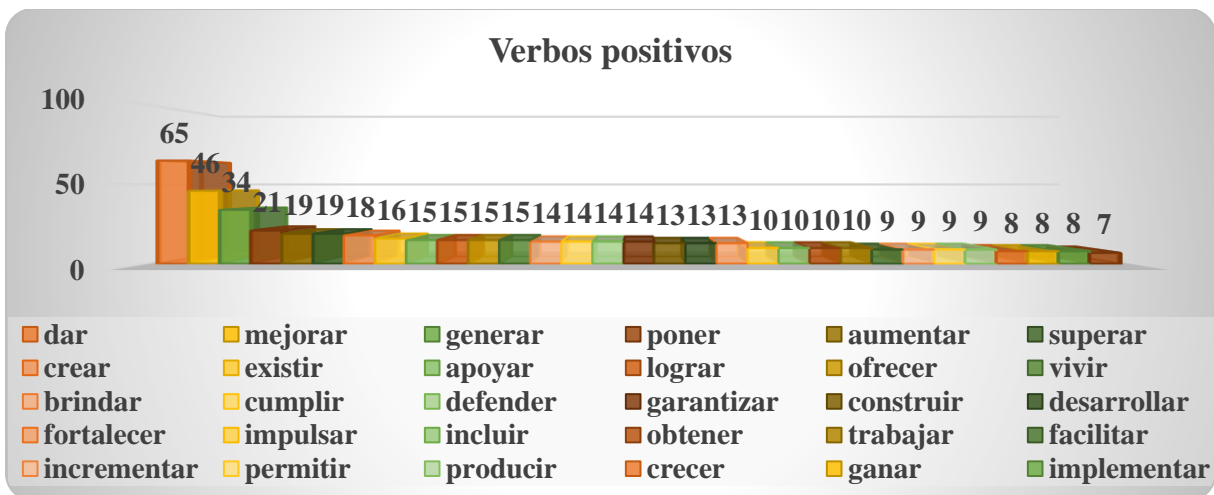


Figura 27: Frecuencia de las ocurrencias según carga del verbo

En los siguientes numerales, se ha descrito cada carga presentando los verbos etiquetados en ellas y listando una serie de ejemplos donde evidenciamos la predilección por las construcciones positivas en los datos analizados.

#### 6.1.7.1. VERBOS POSITIVOS

Con los verbos marcados como positivos, se identificaron 500 ocurrencias correspondientes, como ya se mencionó, al 36,76 %. La figura 28 da cuenta de la distribución de estas ocurrencias en los verbos con esta polaridad. Asimismo, en los numerales 155 a 157, se recuperaron ejemplos del corpus con ocurrencias a partir de verbos positivos.



*Figura 28: Frecuencia de ocurrencias con verbos positivos*

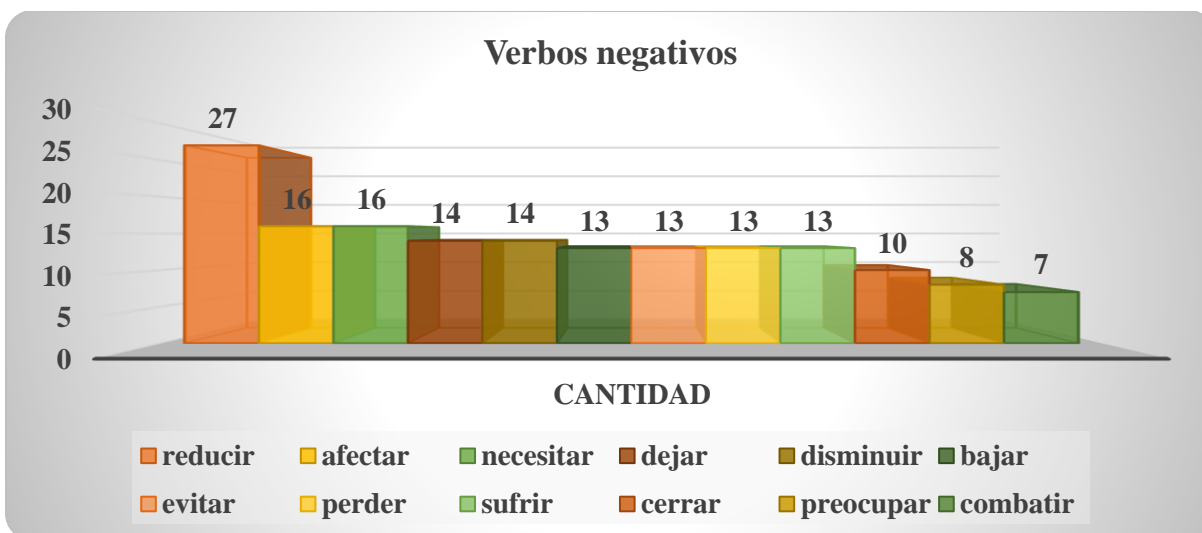
- 155. **dar** vivienda gratis
- 156. **mejorar** calidad de vida
- 157. **generar** enfermedad

Por otro lado, como puede observarse en el numeral 157, un verbo marcado como positivo no necesariamente da como resultado un patrón positivo, situación que también ocurre con los verbos marcados como negativos o neutros. Esta situación se describe más adelante en el numeral sobre las reglas establecidas con las tres polaridades y las posibles cargas finales en el patrón (Capítulo 6 § 6.1.1.1).

### 6.1.7.2. VERBOS NEGATIVOS

Con los verbos negativos, los cuales constituyen el 12,06 % del total de verbos analizados, se construyen más ocurrencias positivas que negativas a la hora de recuperar la polaridad total del patrón; esto es, de las 164 ocurrencias con verbos negativos, 77 son positivas, mientras que 74 son negativas y 14 son neutras. La figura 29 da cuenta de los 12 verbos negativos y la frecuencia de ocurrencias por cada uno. Además, en los numerales 158 a 165 se presentan ejemplos recuperados de los datos, donde se da cuenta de ocurrencias positivas a partir de verbos negativos.



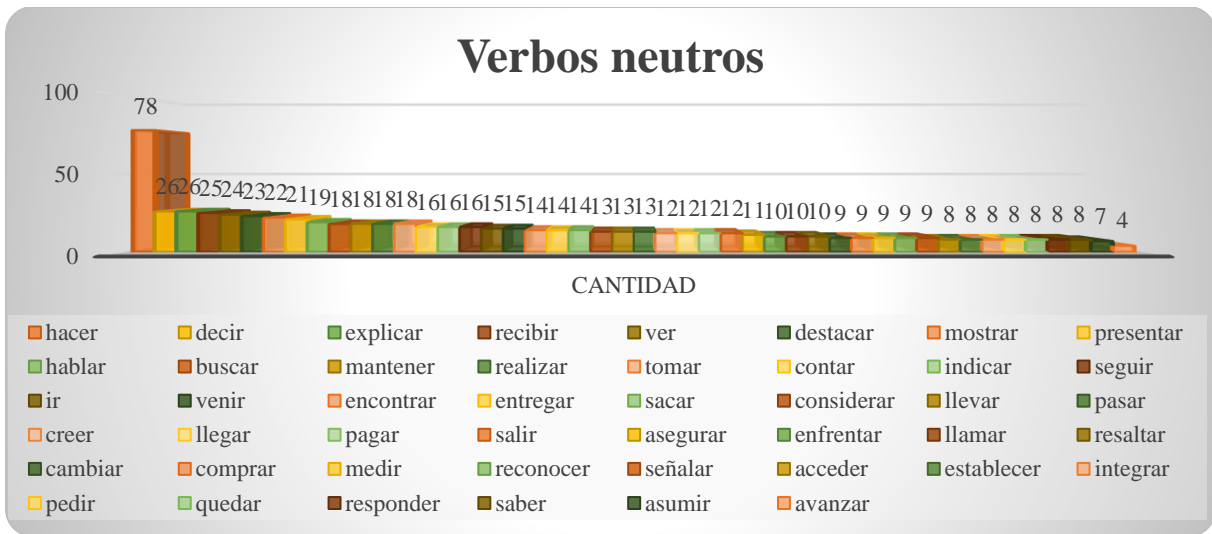


*Figura 29: Frecuencia de ocurrencias con verbos negativos*

- 158. **reducir** pobreza extrema
- 159. **dejar** indigencia
- 160. **disminuir** desigualdad social
- 161. **bajar** costo
- 162. **evitar** abandono
- 163. no **perder** subsidio
- 164. no **sufrir** enfermedad
- 165. **cerrar** brecha de corrupción

### 6.1.7.3. VERBOS NEUTROS

En cuanto a las ocurrencias con verbos neutros, estas son las más frecuentes en los datos analizados, con un porcentaje del 51,69 %. La figura 30 contiene la información de la frecuencia de las ocurrencias de los verbos marcados como neutros, mientras que, en los numerales 166 a 172 se listan ejemplos con verbos neutros.



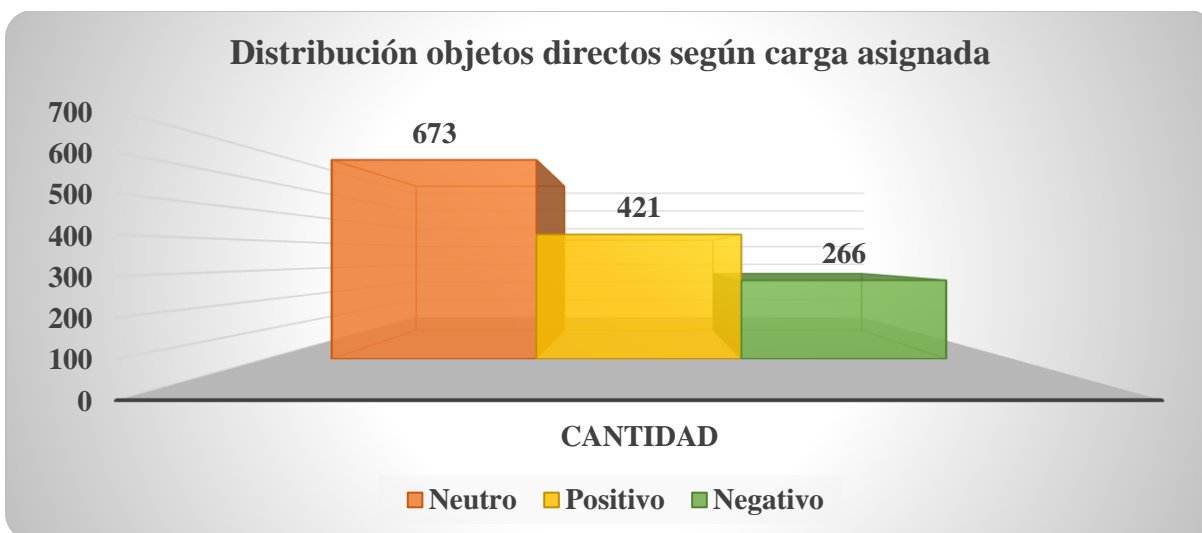
*Figura 30: Frecuencia de ocurrencias con verbos neutros*

- 166. **hacer** pronunciamiento
- 167. **decir** informe de prensa
- 168. **explicar** nuevo acuerdo
- 169. **recibir** sacramento
- 170. **ver** efecto del acuerdo
- 171. **mostrar** oportunidad de crecimiento
- 172. **presentar** asonada

#### 6.1.8. CARGAS DEL OBJETO DIRECTO

A la par del etiquetado de los verbos según las tres polaridades de nuestro trabajo (positivo, negativo o neutro), hemos etiquetado cada uno de los objetos directos, siguiendo el mismo criterio lexicográfico utilizado con los verbos.

En este punto, hemos identificado que se mantiene la regularidad con respecto al uso mayor de ocurrencias positivas o neutras frente a las negativas. El 49,49 % de los objetos directos son neutros, frente al 30,96 % positivos y el 19,56 % negativos. La figura 31 presenta la frecuencia de los objetos directos con relación a la polaridad asignada.



*Figura 31:* Distribución objetos directos según carga asignada

Al igual que con las cargas de polaridad de los verbos, los objetos directos muestran la tendencia a emplear más elementos positivos para hablar de la pobreza, si tenemos en cuenta lo mencionado acerca de lo neutro con tendencia hacia lo positivo. En los numerales 173 a 175 se presentan algunos ejemplos de objetos directos positivos hallados en los datos.

- 173. **Positivo:** agua potable, oportunidad de igualdad, red pública hospitalaria, salud
- 174. **Negativo:** agenda de droga, chip de violencia, minería informal, pobreza extrema
- 175. **Neutro:** informe de prensa, meta establecida, pronunciamiento, programa

Con relación a los objetos directos donde el verbo ha sido nominalizado, hemos constatado que en los casos donde se presenta la ocurrencia nominalizada al igual que la verbal, la polaridad en ambas es la misma.

El análisis de las ocurrencias corresponde a las reglas de combinación de patrones establecidas para nuestros datos, las cuales son explicadas en detalle más adelante (véase Anexo B). En los numerales 176 a 179 se presentan ejemplos para ilustrar este hallazgo, junto a la regla de combinación de la ocurrencia y la carga dada al objeto directo correspondiente a su nominalización.

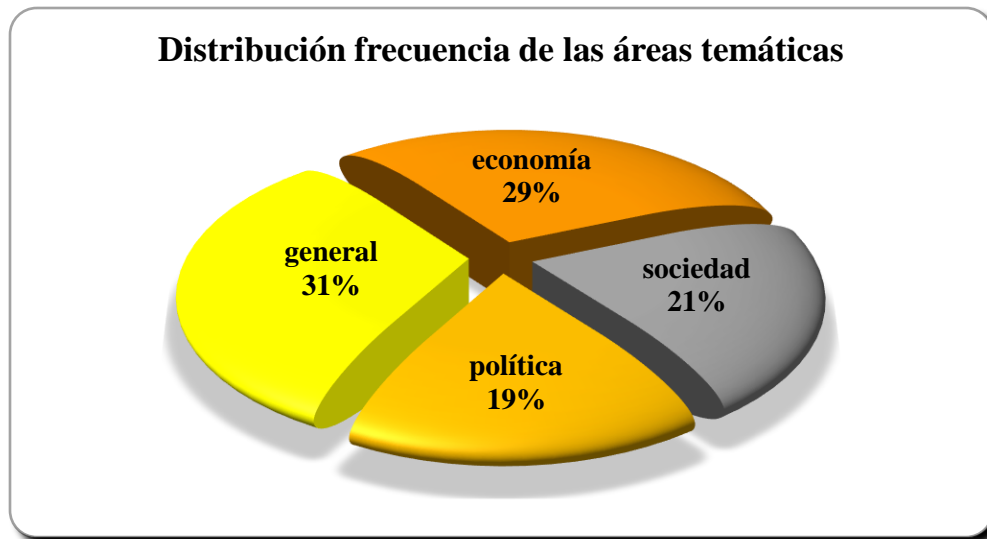
- 176. Reducir pobreza (neg+neg=pos)      reducción de la pobreza      positivo

177.	Crecer economía (pos+neu=pos)	crecimiento económico	positivo
178.	Producir grano (pos+neu=pos)	producción de arroz	neutro
179.	Reducir costo (neg+neu=pos)	reducción de costo	positivo

### 6.1.9. ÁREA TEMÁTICA DE LOS OBJETOS DIRECTOS

Los objetos directos se etiquetaron en 4 áreas temáticas: *general*, *economía*, *política* y *sociedad*, establecidas en la metodología de nuestra investigación (Capítulo 4 § 4.7.2.). Bajo estas cuatro categorías, se agruparon una serie de subcategorías, las cuales fueron halladas en el etiquetaje inicial de los datos. Más adelante, se describe en detalle cómo se construyó cada una de estas áreas temáticas.

Con la asignación de las áreas temáticas a cada uno de los objetos directos, se pudo constatar que las áreas definidas corresponden con las secciones de las cuales se recuperaron los artículos del corpus en los periódicos seleccionados. De igual manera, se pudo identificar que las áreas definidas también corresponden a ámbitos desde los cuales se aborda el tema de la pobreza, a saber, *economía*, *sociedad* y *política*. Con relación a los objetos directos agrupados en la categoría *general*, se ha constatado que estos no poseen marca temática clara. De este modo, su utilización corresponde, con mayor frecuencia, a aportar información de tipo contextual o ser empleados como estrategias discursivas y, en algunos casos, participan en la construcción de la carga de opinión del texto. La figura 32 presenta la distribución de las áreas temáticas de los objetos directos.



*Figura 32:* Distribución frecuencia de las áreas temáticas

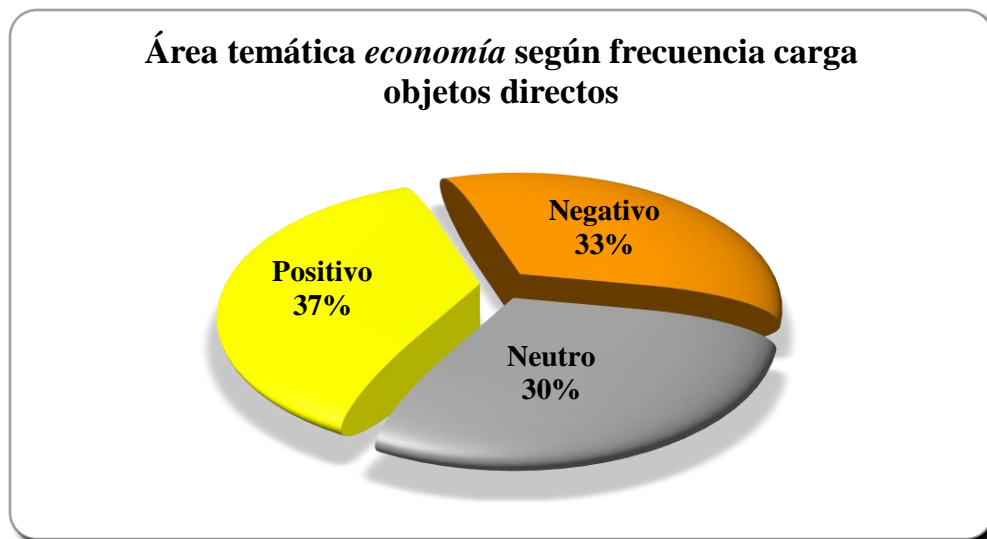
Al contrastar el área temática frente a la carga asignada a los objetos directos, exceptuando el área *economía*, las ocurrencias negativas quedaron relegadas de nuevo a las posiciones finales. En los siguientes numerales, se ha descrito y ejemplificado la frecuencia de cada área temática en contraste con las cargas asignadas a cada objeto directo.

#### **6.1.9.1. ÁREA TEMÁTICA *ECONOMÍA* SEGÚN DISTRIBUCIÓN DE LAS CARGAS DE LOS OBJETOS DIRECTOS**

En *economía* se agruparon las ocurrencias relativas al empleo, desarrollo de los sectores económicos, inversiones por parte de empresas privadas o públicas, índices con relación a los datos de pobreza, explotación o producción de bienes, oferta de servicios, explotación de recursos y demás ítems relacionados con dinero como el tema de regalías, pensiones, presupuestos, entre otros.

Respecto a la distribución de las ocurrencias según la carga asignada en *economía*, la tendencia que se ha planteado se mantiene, al ser encontrados, en mayor proporción, objetos directos positivos frente a los negativos o neutros.

La figura 33 presenta la distribución de los objetos directos en *economía* según las cargas asignadas.



*Figura 33: Área temática *economía* según las cargas de los objetos directos*

Al establecer la relación directa entre el tema trabajado (pobreza) y esta área temática (*economía*), se ha descubierto, una vez más, que la dinámica tiende a presentar una mejoría de la problemática, en cuanto los índices de pobreza o se reducen o se mejoran los indicadores que determinan los criterios de estos conforme al IMP, como se ilustra con objetos directos positivos en los numerales 180 a 182; negativos en los numerales 183 a 185, y neutros en los numerales 186 a 188.

- 180. mantener **desarrollo agrícola**
- 181. permitir **distribución equitativa**
- 182. aumentar **empleo permanente**
  
- 183. dejar **pobreza extrema**
- 184. disminuir **tasa de desempleo**
- 185. reducir **hambre**
  
- 186. desarrollar **campo**
- 187. pagar **cuota**

188. realizar **compra**

### 6.1.9.2. **ÁREA TEMÁTICA *SOCIEDAD* SEGÚN DISTRIBUCIÓN DE LAS CARGAS DE LOS OBJETOS DIRECTOS**

En *sociedad* se agruparon las ocurrencias correspondientes a educación, tecnología, construcción, medicina, medio ambiente, cultura, deporte, medios de transporte y comunicación, religión, entre otras.

Básicamente, las ocurrencias agrupadas remiten a la cotidianidad. De nuevo, las ocurrencias negativas son relegadas a la última casilla según la distribución de estas conforme a la carga asignada, información presentada en la figura 34.

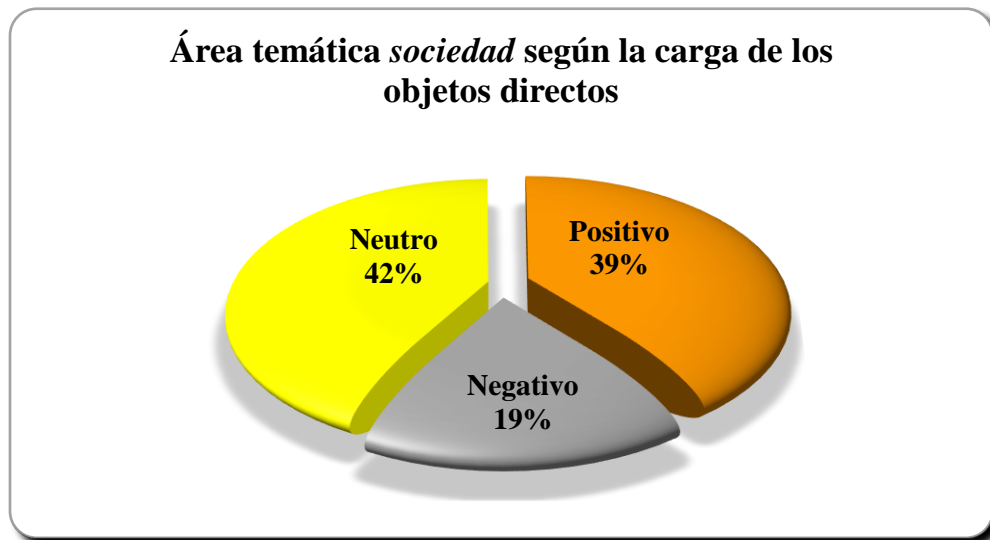


Figura 34: Área temática *sociedad* según la carga de los objetos directos

Al igual que *economía*, por medio de esta área temática, se ha descubierto que se presenta, principalmente, una visión positiva en el bienestar de las personas. Los ejemplos listados en los numerales 189 a 197, recuperados de los datos y presentados en series de tres con objetos directos positivos, negativos y neutros, respectivamente, dan cuenta de lo aquí planteado.

189. acceder **seguridad social**

190. mantener **dinámica**

191. tomar **curso**

- 192. cambiar **agenda de droga**
- 193. presentar **asonada**
- 194. ver **acaparamiento**
  
- 195. avanzar **movilidad social**
- 196. cumplir **criterio ambiental básico**
- 197. facilitar **permiso de construcción**

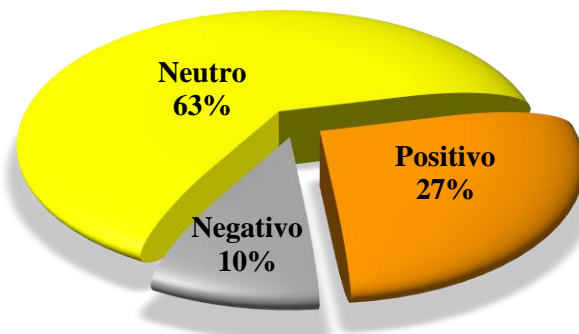
### **6.1.9.3. ÁREA TEMÁTICA *POLÍTICA* SEGÚN DISTRIBUCIÓN DE LAS CARGAS DE LOS OBJETOS DIRECTOS**

En *política*, se agruparon las ocurrencias correspondientes a información estatal desde los ámbitos judicial, legal y administrativo del tipo tratados, leyes, proyectos, ámbito electoral, militar y política internacional, entre otros.

Estas ocurrencias dan cuenta del manejo político de la pobreza y las aristas asociadas con relación a las distintas estrategias desde lo estatal y/o gubernamental para la consecución de los objetivos con relación a la erradicación de la pobreza (ONU, 2015). La figura 35 presenta la distribución de las cargas de los objetos directos en esta área. En los numerales 198 a 200, se listan ejemplos de ocurrencias con objetos directos con carga neutra; en los numerales 201 a 203, con carga positiva y en los numerales 204 a 206, con carga negativa.



**Área temática *política* según carga de los objetos directos**



*Figura 35: Área temática *política* según cargas de los objetos directos*

- 198. apoyar **proyecto europeo**
- 199. crear **política pública masiva**
- 200. mejorar **condición**
  
- 201. aprobar **ley**
- 202. construir **vivienda social**
- 203. defender **libertad de prensa**
  
- 204. combatir **narcotráfico**
- 205. dejar **guerra**
- 206. evitar **paro**

En cuanto al carácter positivo (unido al neutro) de las ocurrencias de *política* (que son las más frecuentes), podría establecerse que las estrategias empleadas desde los entes de gobierno, han dado sus resultados al ser referidos desde una visión positiva. Las estrategias planteadas desde los distintos estamentos de gobierno en Colombia con miras a la erradicación de la pobreza, según los

datos analizados, han sido acertados en cuanto demuestran la reducción del fenómeno como se muestra en los ejemplos de los numerales 207 a 209.

- 207.       aumentar seguridad
- 208.       desarrollar sistema de transporte
- 209.       garantizar inclusión

#### 6.1.9.4.   **ÁREA TEMÁTICA GENERAL SEGÚN DISTRIBUCIÓN DE LAS CARGAS DE LOS OBJETOS DIRECTOS**

Por último, bajo el área temática *general* se agruparon las ocurrencias cuya marca temática era nula o poco clara. Gran parte de estas ocurrencias son de uso discursivo en cuanto pueden introducir o contextualizar las ocurrencias agrupadas en las tres categorías anteriores.

Se presenta una mayor frecuencia de las ocurrencias neutras, seguidas de las positivas y, por último, las negativas. La figura 36 ilustra esta distribución. En los numerales 210 a 212 se listaron ejemplos de ocurrencias en el área temática general con objetos directos con carga neutra; en los numerales 213 a 215, con carga positiva y en los numerales 216 a 218, con carga negativa.

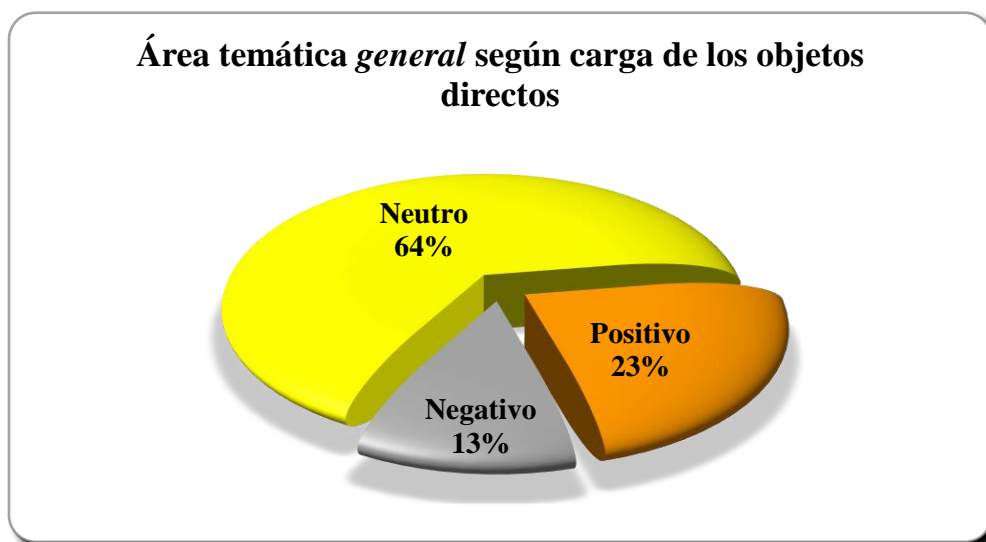


Figura 36: Área temática *general* según cargas de los objetos directos

- 210.       dar **herramienta**

- 211. explicar **informe**
- 212. ofrecer **información**
  
- 213. aumentar **cobertura**
- 214. lograr **mejora**
- 215. sacar **provecho**
  
- 216. enfrentar **dificultad**
- 217. presentar **disturbio**
- 218. responder **ataque**

#### 6.1.10. CARGAS OBJETOS DIRECTOS CON RELACIÓN AL ÁREA TEMÁTICA

Después de describir cómo se construyó cada área temática y de puntualizar el tipo de información aportada en cada una, la relación entre área temática y la carga del argumento, nos aporta otro elemento en la identificación de las regularidades en cuanto a la polaridad de las ocurrencias que, hasta el momento, hablan de una predilección por lo positivo y lo neutro frente a lo negativo. En la tabla 27 se presenta la distribución de los objetos directos por área temática.

Tabla 27: Distribución objetos directos según área temática

<b>Distribución objetos directos según área temática</b>		
<b>Área temática</b>	<b>Frecuencia</b>	<b>%</b>
<b>general</b>	425	31,25
<b>economía</b>	397	29,19
<b>sociedad</b>	279	20,51
<b>política</b>	259	19,04
Palabras o elementos distintos contados: 4		
<b>Total:</b>	1.360	100 %

Teniendo claro que en la categoría *general* se agrupan los objetos directos sin marca temática clara, se relega la función de establecer temáticamente el fenómeno en las tres categorías restantes. El orden hallado en la distribución de los objetos directos, según la frecuencia en estas áreas, podría

establecer cómo se aborda la pobreza desde lo económico (*economía*), luego desde el sujeto (*sociedad*) y, en última instancia, desde la autoridad representada en la institucionalidad (*política*).

Desde la perspectiva semiótica de Pardo y Ruiz (2017) y la sociológica de Álvarez y Naharro (2017), las cuales retomamos para nuestro trabajo, se ha planteado que la pobreza ha sido abordada, principalmente, desde un ámbito económico aunque, en sus trabajos, estos autores proponen ver la pobreza desde un enfoque más integral y no solo en términos de posesión.

Respecto a la distribución de los objetos directos con relación al área temática, es factible corroborar lo expresado desde estos autores, en cuanto al tratamiento predominante de la pobreza desde aspectos económicos. En los siguientes numerales, se presentan los resultados con la relación entre la frecuencia de las cargas de los objetos directos y el área temática.

#### 6.1.10.1. OBJETOS DIRECTOS POSITIVOS CON RELACIÓN AL ÁREA TEMÁTICA

Un total de 425 ocurrencias corresponden a los objetos directos positivos en las 4 áreas temáticas. La figura 37 relaciona la cantidad de ocurrencias por área temática, donde *economía* es la más frecuente con un 34,44 %, mientras que *política* es la de menor frecuencia con un 16,63 %. En los numerales 219 a 222, se listan ejemplos de objetos directos positivos con relación al área temática, recuperados de nuestros datos.

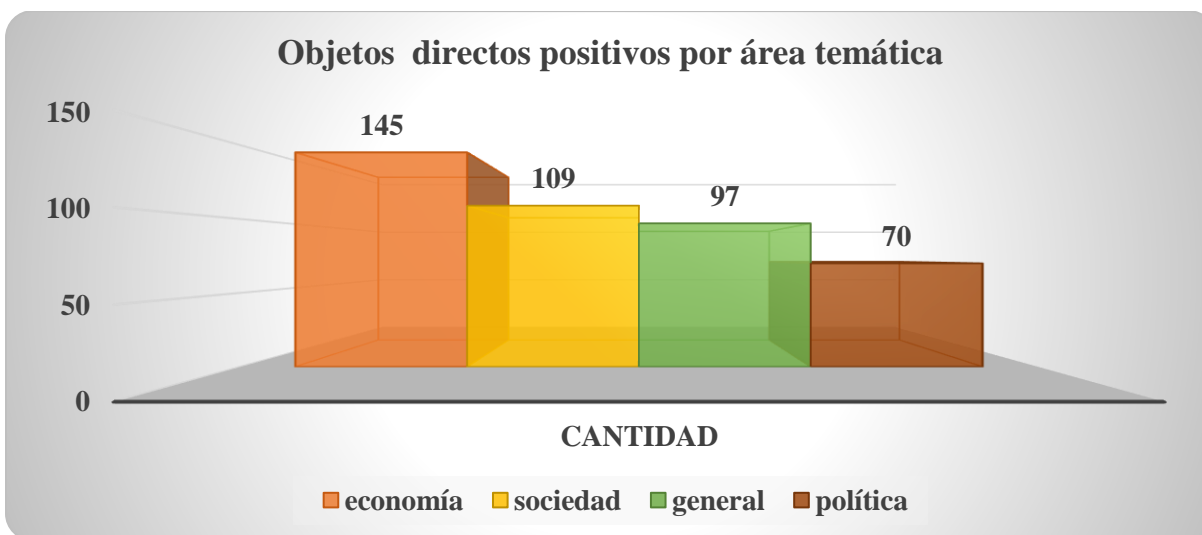


Figura 37: Objetos directos positivos por área temática

219. *Economía*: considerar **ganancia**, crear **empleo**, dar **trabajo**, destacar **inversión**

- 220. *Sociedad*: construir **hospital**, contar **agua potable**, mejorar **calidad de vida**
- 221. *General*: existir **igualdad**, hacer **avance**, recibir **beneficio**, recibir **premio**
- 222. *Política*: apoyar **gobierno colombiano**, aprobar **ley**, defender **proceso de paz**

### 6.1.10.2. OBJETOS DIRECTOS NEGATIVOS CON RELACIÓN AL ÁREA TEMÁTICA

En el caso de los objetos directos negativos, el área de *economía* es igualmente la más frecuente con un 49,25 % frente a la *política* con un 9,77 %, siendo esta última la menos frecuente. La figura 38 corresponde a la distribución de las ocurrencias con objetos directos negativos. En los numerales 223 a 226 se muestran ejemplos de objetos directos negativos con relación al área temática.

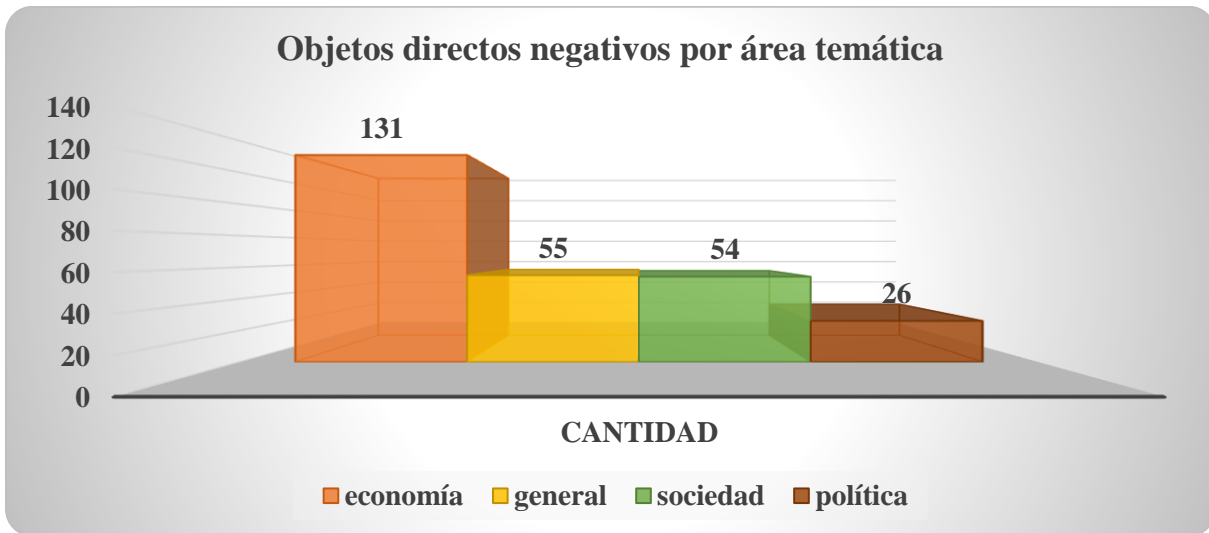


Figura 38: Objetos directos negativos por área temática

- 223. *Economía*: bajar **intensidad de la pobreza**, sacar **pobreza extrema**, seguir **brecha**
- 224. *General*: enfrentar **problema estructural**, explicar **dificultad**, producir **tristeza**
- 225. *Sociedad*: evitar **abandono**, disminuir **muerte**, generar **enfermedad**, pasar **arma**
- 226. *Política*: combatir **corrupción**, crear **violencia**, mantener **destierro**, vivir **guerra**

### 6.1.10.3. OBJETOS DIRECTOS NEUTROS CON RELACIÓN AL ÁREA TEMÁTICA

En última instancia, los objetos directos neutros corresponden al 49 % de los datos. Contrario a los objetos directos positivos y negativos, el área temática más frecuente, para el caso de los objetos directos neutros, es el área *general* con un porcentaje del 40,56 % frente al 17,24 % de las ocurrencias en *sociedad*. La figura 39 contiene la distribución de los objetos directos neutros por área temática. De nuevo, se presentan ejemplos sobre objetos directos neutros en cada área temática, recuperados de los datos en los numerales 227 a 230.

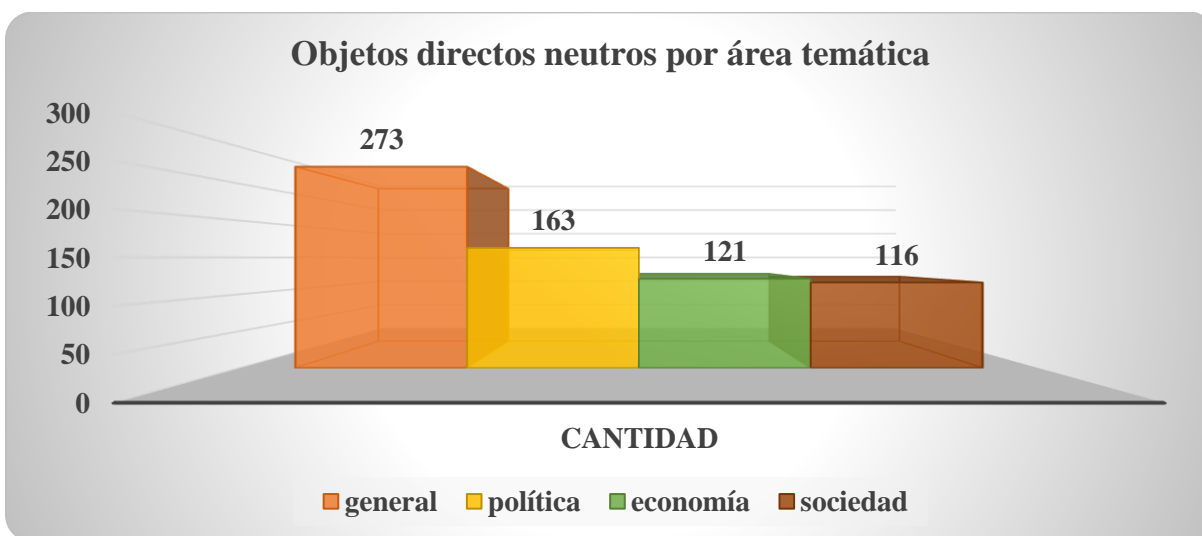


Figura 39: Objetos directos neutros por área temática

- 227. *General*: asegurar **cambio de metodología**, llamar **atención**, lograr **cambio**
- 228. *Política*: buscar **senda**, cumplir **asistencia**, decir **Cepal**, seguir **espacio político**
- 229. *Sociedad*: destacar **papel educación**, realizar **cambio estructural**, tomar **curso**
- 230. *Economía*: dar **cierre financiero**, pagar **monto**, realizar **compra**, reducir **costo**

### 6.1.11. CARGA TOTAL PATRÓN [VERBO + OBJETO DIRECTO]

En la metodología, la fase de asignación de la polaridad concluyó con la polaridad del patrón, es decir, la suma de la polaridad del verbo más la polaridad del objeto directo. Es importante mencionar que, para los fines de nuestra investigación, el peso asignado a los elementos corresponde a una variable discreta y no continua, como es común en bases de datos de este tipo

(Véase *WordNet* (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990), *SenSem* (Alonso, Capilla, Castellón, Fernández-Montraveta, & Vázquez, 2005), *Senticon* (Cruz, Troyano, Pontes, & Ortega, 2014), entre otras). La asignación de la carga total del patrón se realizó de forma manual y se validó por los dos lingüistas expertos.

Al tener las ocurrencias con sus respectivas cargas, a saber, carga verbo, objeto directo y patrón, se identificaron las posibles reglas de combinación con las tres polaridades de nuestro trabajo (Véase Anexo B). Se definieron las siguientes siglas para las polaridades: POS positivo, NEG negativo y NEU neutro. El orden de los patrones es el siguiente: [verbo con carga X + objeto directo con carga Y = carga final del patrón].

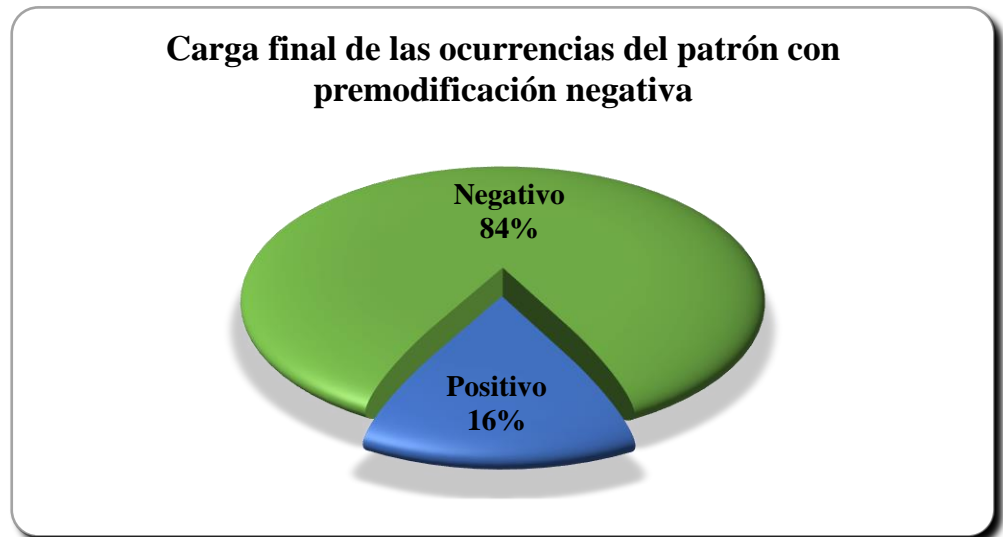
Es necesario tener en cuenta que 43 ocurrencias cuentan con premodificación negativa (3,16 %), puesto que, para la asignación de la carga del patrón final, la premodificación negativa se convierte en un conmutador de polaridad; es decir, la carga final asignada a la ocurrencia se invierte en aquellas que cuentan con premodificación negativa. De allí, que se originen combinaciones del tipo [POS+POS=NEG], [POS+NEG=POS], [POS+NEU=NEG], entre otras. En los numerales 231 a 235 se presentan algunos ejemplos de ocurrencias con premodificación negativa.

- 231. (no) brindar gran beneficio
- 232. (no) dar subsidio
- 233. (no) generar calidad de vida
- 234. (no) crear impuesto
- 235. (no) tomar medida radical

Al momento de conmutar la carga de los patrones con dicha premodificación, hemos hallado que estos tienden a tomar una carga final negativa por encima de la positiva y que no se presentan casos de patrones premodificados cuya carga final sea neutra. Asimismo, con base en los datos analizados, la polaridad solo se conmutó en patrones con carga inicial positiva o negativa.

La figura 40 corresponde a la carga final asignada a las ocurrencias del patrón con premodificación negativa luego de ser conmutada la carga inicial. En los numerales 236 a 239 se

presentan algunos ejemplos de patrones con premodificación negativa luego de conmutar la carga inicial.

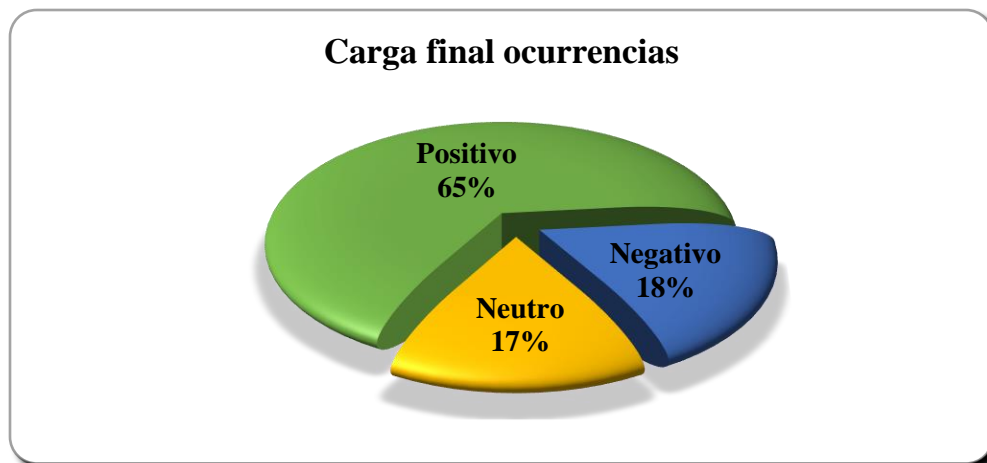


*Figura 40:* Carga final de las ocurrencias del patrón con premodificación negativa

- 236. (no) recibir pensión – carga final negativa
- 237. (no) encontrar solución - carga final negativa
- 238. (no) sufrir enfermedad - carga final positiva
- 239. (no) quedar cansancio - carga final positiva

La figura 41 corresponde a la distribución de la frecuencia de las cargas finales de las ocurrencias del patrón luego de conmutar las ocurrencias con premodificación.





*Figura 41: Frecuencia carga final ocurrencias*

Las ocurrencias positivas corresponden al 65 % de los datos (890), mientras que las negativas dan cuenta del 18 % de los datos (244). Por último, el 17 % son ocurrencias neutras (226). En los numerales 240 a 242 se listan algunos ejemplos de ocurrencias positivas; en los numerales 243 a 245, las negativas y, en los numerales 246 a 248, las neutras.

- 240. reducir barrera comercial
- 241. brindar oportunidad de igualdad
- 242. mejorar calidad de la educación
  
- 243. indicar hambre
- 244. presentar nivel de pobreza
- 245. perder empleo
  
- 246. dar indicador
- 247. hablar prensa
- 248. explicar bondad

Se conserva la tendencia a encontrar ocurrencias positivas (unidas a la neutras) con mayor frecuencia que las negativas. Es decir, de las 1.360 ocurrencias analizadas, el 82 % hacen

referencia a una visión positiva de la pobreza mientras el 18 % restante corresponden a la visión negativa. Se podría interpretar que la pobreza en Colombia tiende a ser una problemática cuya representación en la prensa es principalmente positiva.

Se ha encontrado en nuestro corpus ocurrencias que cumplen con 26 de las 27 posibles reglas de combinación; esto es, la única posibilidad de combinación que no se encontró en los datos, fue POS+POS=NEU. Se presenta una lista con las 26 reglas de combinación junto a una serie de ejemplos recuperados de los datos (Véase el Anexo C).

En la figura 42 se presenta la frecuencia de los patrones conforme a las posibles combinaciones, de acuerdo con las reglas recuperadas de los datos analizados.

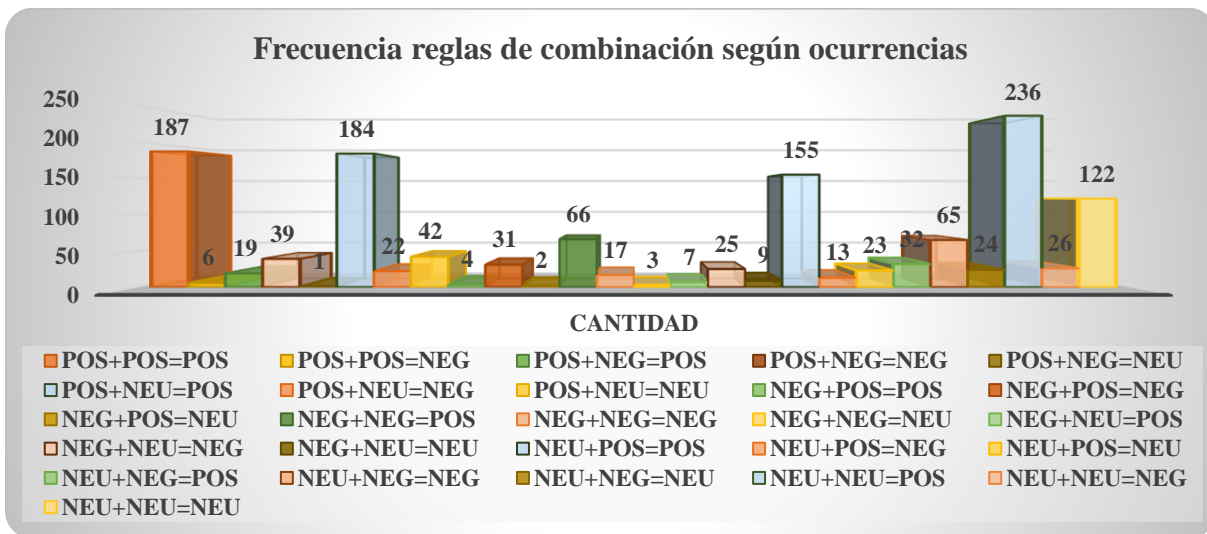


Figura 42: Frecuencia reglas de combinación según ocurrencias

Con la información aportada por la carga total de las ocurrencias analizadas, se ha ratificado la tendencia positiva sobre la pobreza en los periódicos analizados. Se podría establecer que la prevalencia de las ocurrencias neutras y positivas dan cuenta de un discurso en el cual la pobreza en Colombia es un fenómeno que, aunque existente, su dinámica indica que se ha ido superando, al menos en el tiempo analizado (2010-2014).

En los siguientes numerales, se presentan los resultados obtenidos a la hora de contrastar la carga final de las ocurrencias con los demás elementos tenidos en cuenta en el análisis (tipos de verbos, sentidos de verbos y las áreas temáticas de los objetos directos). Estas categorías ya han

sido descritas previamente. En consecuencia, la distribución de las cargas, en cada ítem, da fundamento a las ideas expresadas con relación a los hallazgos descritos. En general, se presentan las frecuencias para cada categoría, al igual que ejemplos recuperados de los datos en consonancia con la dinámica de análisis hasta ahora seguida.

### **6.1.12. CARGA TOTAL PATRÓN [VERBO + OBJETO DIRECTO] CON RELACIÓN AL TIPO DEL VERBO**

Con el tipo de verbo damos cuenta de la elección léxica en la presentación de la pobreza en la prensa en Colombia. Con los verbos fraseológicos y conectores, se establece el nivel de especialidad del tema abordado, mientras que con los verbos discursivos se describen las estrategias en la presentación del fenómeno. A continuación, se ilustra y se ejemplifica el comportamiento de las ocurrencias según el tipo de verbo con relación a la polaridad.

#### **6.1.12.1. POSITIVO**

La distribución de las 890 ocurrencias positivas se presenta en la tabla 28. Los numerales 249 a 251 corresponden a ejemplos de ocurrencias positivas según el tipo de verbo.

*Tabla 28:* Tipo de verbo según carga positiva del patrón

<b>Tipo de verbo según carga positiva del patrón</b>		
<b>Tipo de verbo</b>	<b>Frecuencia</b>	<b>%</b>
Fraseológico	589	66,18
Discursivo	173	19,44
Conector	128	14,38
Palabras o elementos distintos contados: 3		
<b>Total:</b>	<b>890</b>	<b>100</b>

249. *Fraseológicos:* cerrar brecha, construir vivienda social, incrementar comercio

250. *Discursivo:* destacar papel educación, explicar reducción de la pobreza

251. *Conector:* dar acceso, dar vivienda gratis, entregar arma, incluir servicio

Al hablar de las cargas de los verbos con relación al tipo, se ha concluido que los verbos fraseológicos son aquellos discriminantes con relación a la polaridad. Los resultados aquí

expuestos, con relación a la carga final positiva de las ocurrencias, confirman y validan nuestra afirmación.

### 6.1.12.2. NEGATIVO

En la tabla 29 se presenta la distribución de las 244 ocurrencias negativas. En los numerales 252 a 254 se listan ejemplos conforme a cada tipo de verbo.

Tabla 29: Tipo de verbo según carga negativa del patrón

Tipo de verbo según carga negativa del patrón		
Tipo de verbo	Frecuencia	%
Fraseológico	158	64,75
Conector	44	18,03
Discursivo	42	17,21
Palabras o elementos distintos contados: 3		
Total:	244	100

- 252. *Fraseológico*: aumentar desigualdad, mantener austeridad, sufrir millonaria perdida
- 253. *Conector*: afectar condición de vida, encontrar consumo de sustancia psicoactiva
- 254. *Discursivo*: considerar índice de miseria, ir pobre, reconocer crisis de la salud

En concordancia con lo expresado sobre los verbos fraseológicos como discriminantes de la polaridad, la tendencia se mantiene y a la vez se ratifica con los resultados obtenidos, teniendo las ocurrencias finales con carga negativa como elemento de análisis.

### 6.1.12.3. NEUTRO

Por último, la tabla 30 reúne los datos con relación a las 226 ocurrencias con carga neutra. En los numerales 255 a 257 se listan los ejemplos correspondientes.

Tabla 30: Tipo de verbo según carga neutra del patrón

<b>Tipo de verbo según carga neutra del patrón</b>		
<b>Tipo de verbo</b>	<b>Frecuencia</b>	<b>%</b>
Discursivo	109	48,23
Fraseológico	80	35,4
Conector	37	16,37
Palabras o elementos distintos contados: 3		
Total:	226	100

255. *Discursivo*: dar indicadores de población, indicar modelo de desarrollo rural

256. *Fraseológico*: buscar sector privado, considerar evolución de la pobreza

257. *Conector*: dar declaración conjunta, preocupar tema, desarrollar proyecto

El hecho de que los verbos discursivos ocupen el primer lugar en la distribución de las ocurrencias neutras, ratifica la función de esta polaridad en la construcción de los textos a partir de las distintas estrategias discursivas utilizadas por el autor. Los resultados aquí presentados, concuerdan con las características que hemos definido tanto para los tipos de verbo como para las cargas de polaridad.

### **6.1.13. CARGA TOTAL PATRÓN [VERBO + OBJETO DIRECTO] CON RELACIÓN AL SENTIDO DEL VERBO**

El sentido del verbo da cuenta de la elección del vocabulario con relación a los criterios del IMP y la forma como estos se abordan en la representación de la pobreza en los periódicos elegidos. En los siguientes numerales, se ha descrito el comportamiento de los sentidos del verbo según la carga de la ocurrencia. Se presenta la distribución de las ocurrencias por carga y sentido de verbo junto a ejemplos para ilustrar cada caso.

#### **6.1.13.1. POSITIVO**

De las 890 ocurrencias positivas, los verbos con mayor frecuencia son aquellos de *possession* seguidos por verbos de *change*, *social*, *communication*, *creation*, *motion*, *cognition*, *perception* y *stative* (estos corresponden a los sentidos con una frecuencia superior al 5 %). La tabla 31 presenta la frecuencia por sentido de cada verbo.

Tabla 31: Sentido del verbo según carga positiva del patrón

<b>Sentido del verbo según carga positiva del patrón</b>		
<b>Sentido</b>	<b>Frecuencia</b>	<b>%</b>
<i>possession</i>	168	18,88
<i>change</i>	145	16,29
<i>social</i>	136	15,28
<i>communication</i>	116	13,03
<i>creation</i>	72	8,09
<i>motion</i>	60	6,74
<i>cognition</i>	48	5,39
<i>perception</i>	47	5,28
<i>stative</i>	45	5,06
<i>competition</i>	32	3,6
<i>contact</i>	17	1,91
<i>consumption</i>	2	0,22
<i>body</i>	1	0,11
<i>emotion</i>	1	0,11
Palabras o elementos distintos contados: 14		
<b>Total:</b>	<b>890</b>	<b>100</b>

En los numerales 258 a 266 se presentan ejemplos de ocurrencias positivas de acuerdo con los sentidos cuya frecuencia es superior al 5 %.

- 258. comprar alimento (*possession*)
- 259. cambiar chip de violencia (*change*)
- 260. hacer redistribución de la riqueza (*social*)
- 261. enfrentar rezago (*communication*)
- 262. realizar compra (*creation*)
- 263. seguir cobertura (*motion*)
- 264. establecer ley (*cognition*)
- 265. señalar crecimiento económico (*perception*)
- 266. mantener desarrollo agrícola (*stative*)

### 6.1.13.2. NEGATIVO

De las 244 ocurrencias positivas, los verbos con mayor frecuencia son aquellos de *change*, seguidos por verbos de *possession*, *stative*, *perception*, *cognition*, *social*, *creation*, y *motion*. La tabla 32 presenta la frecuencia por sentido de cada verbo; sin embargo, solo se abordan en detalle los sentidos con una frecuencia superior al 5 %.

Tabla 32: Sentido del verbo según carga negativa del patrón

Sentido del verbo según carga negativa del patrón		
Sentido	Frecuencia	%
<i>change</i>	49	20,08
<i>possession</i>	44	18,03
<i>stative</i>	35	14,34
<i>perception</i>	28	11,48
<i>cognition</i>	21	8,61
<i>social</i>	19	7,79
<i>creation</i>	13	5,33
<i>motion</i>	13	5,33
<i>communication</i>	12	4,92
<i>competition</i>	5	2,05
<i>body</i>	4	1,64
<i>contact</i>	1	0,41
Palabras o elementos distintos contados: 12		
Total:	244	100

En los numerales 267 a 274 se presentan ejemplos de ocurrencias negativas de acuerdo con los sentidos cuya frecuencia es superior al 5 %.

- 267. integrar banda (*change*)
- 268. contar redes de saneamiento (*possession*)
- 269. quedar resentimiento (*stative*)
- 270. ver acaparamiento (*perception*)
- 271. vivir desempleo (*cognition*)
- 272. hacer mella (*social*)

273. (no) generar calidad de vida (*creation*)
274. seguir brecha (*motion*)

### 6.1.13.3. NEUTRO

Las 226 ocurrencias neutras se distribuyen en un total de 12 sentidos, siendo los verbos de *communication* los más recurrente, seguidos de los verbos de *possession*, *perception*, *creation*, *change*, *cognition* y *social* (estos con frecuencia superior al 5 %). La tabla 33 presenta la frecuencia por sentido de cada verbo.

Tabla 33: Sentido del verbo según carga neutra del patrón

<b>Sentido del verbo según carga neutra del patrón</b>		
<b>Sentido</b>	<b>Frecuencia</b>	<b>%</b>
<i>communication</i>	86	38,05
<i>possession</i>	23	10,18
<i>perception</i>	21	9,29
<i>creation</i>	19	8,41
<i>change</i>	16	7,08
<i>social</i>	16	7,08
<i>cognition</i>	15	6,64
<i>motion</i>	10	4,42
<i>stative</i>	9	3,98
<i>emotion</i>	7	3,1
<i>contact</i>	3	1,33
<i>competition</i>	1	0,44
Palabras o elementos distintos contados: 12		
<b>Total:</b>	<b>226</b>	<b>100</b>

En los numerales 275 a 281 se presentan ejemplos de ocurrencias neutras de acuerdo con los sentidos cuya frecuencia es superior al 5 %.

275. explicar informe (*communication*)
276. recibir llamada (*possession*)
277. señalar política de vivienda (*perception*)
278. generar expectativa (*creation*)



- 279. sacar cigarrillo (*change*)
- 280. medir pobreza (*cognition*)
- 281. hacer comentario (*social*)

#### 6.1.14. CARGA TOTAL PATRÓN [VERBO + OBJETO DIRECTO] CON RELACIÓN AL ÁREA TEMÁTICA

El área temática del objeto directo determina los ejes desde los cuales se presenta la pobreza en los periódicos. A excepción de la polaridad neutra, en su mayoría compuesta por elementos discursivos y las ocurrencias en la categoría *general*, las áreas temáticas se abordan en el mismo orden: *economía*, *sociedad* y *política*, lo que establecería, como ya lo hemos mencionado, la jerarquía a la hora de presentar la pobreza. Por último, en los siguientes numerales, se ofrece información acerca de las polaridades *positiva*, *negativa* y *neutra* con relación a la frecuencia de las áreas temáticas y ejemplos ilustrativos.

##### 6.1.14.1. POSITIVO

En la figura 43 se establece la distribución de las áreas temáticas en las ocurrencias positivas. La *economía* aparece en primer lugar con 277 ocurrencias frente a 256 de *general*, 184 de *sociedad* y 173 de *política*.

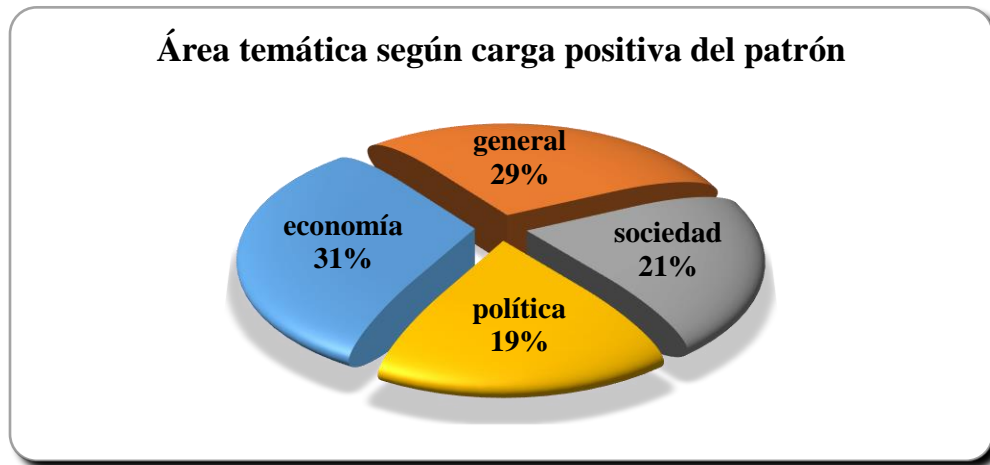


Figura 43: Área temática del objeto directo según carga positiva del patrón

En los numerales 282 a 293 se presentan ejemplos de ocurrencias positivas según las áreas temáticas.

- 282. dejar pobreza (*economía*)
- 283. (no) perder subsidio (*economía*)
- 284. producir riqueza (*economía*)
  
- 285. garantizar transparencia (*general*)
- 286. lograr intervención (*general*)
- 287. resaltar cobertura (*general*)
  
- 288. cambiar vida (*sociedad*)
- 289. garantizar comunicación (*sociedad*)
- 290. mejorar respiración (*sociedad*)
  
- 291. aprobar Estatuto Anticorrupción (*política*)
- 292. buscar reelección (*política*)
- 293. establecer ley (*política*)

#### **6.1.14.2. NEGATIVO**

En la figura 44 se establece la distribución de las áreas temáticas en las ocurrencias negativas. El área temática *economía* aparece en primer lugar con 75 ocurrencias frente a 72 de *general*, 61 de *sociedad* y 36 de *política*.

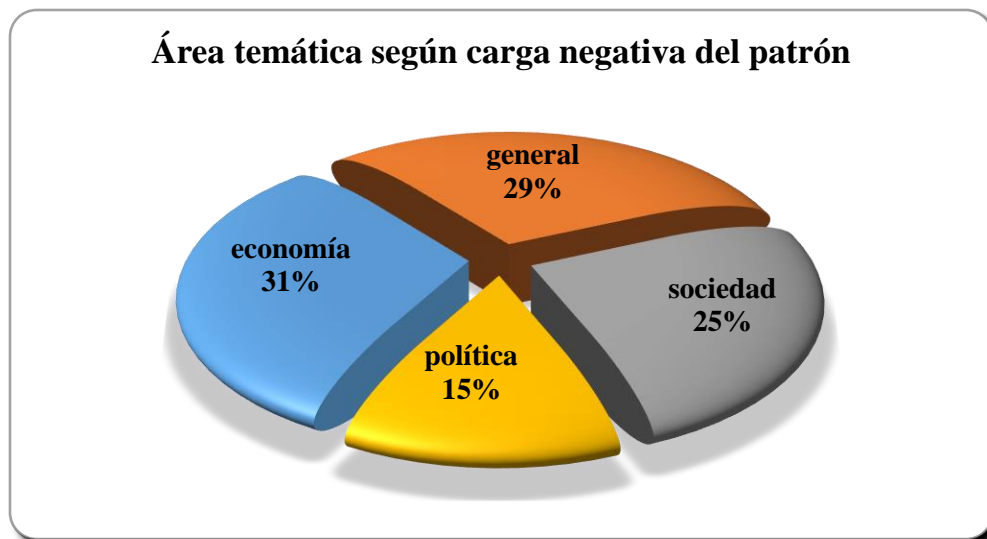


Figura 44: Área temática del objeto directo según carga negativa de la ocurrencia

En los numerales 294 a 305 se presentan ejemplos de ocurrencias negativas según las áreas temáticas.

- 294. aumentar desigualdad (*economía*)
- 295. sufrir millonaria perdida (*economía*)
- 296. vivir desempleo (*economía*)
  
- 297. (no) dar resultado (*general*)
- 298. necesitar cuidado (*general*)
- 299. venir incertidumbre (*general*)
  
- 300. (no) contar servicio de energía (*sociedad*)
- 301. (no) mejorar calidad de vida (*sociedad*)
- 302. perder esperanza (*sociedad*)
  
- 303. afectar escándalo de espionaje (*política*)
- 304. buscar apalancamiento (*política*)

305. perder mar (*política*)

### 6.1.14.3. NEUTRO

En la figura 45 se establece la distribución de las áreas temáticas en las ocurrencias etiquetadas como neutras. El área *general* aparece en primer lugar con 97 ocurrencias frente a 50 de *política*, 45 de *economía* y 34 de *sociedad*.

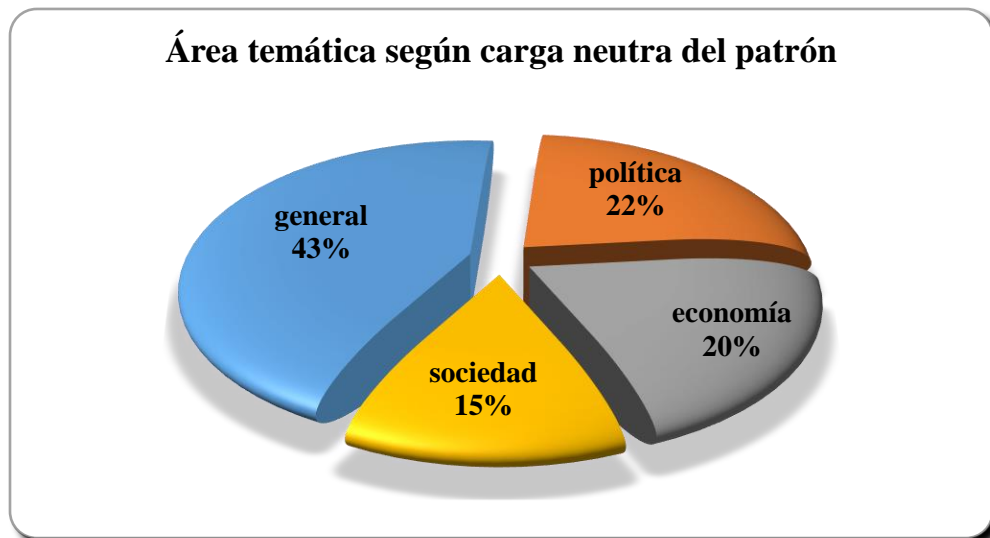


Figura 45: Área temática del objeto directo según carga neutra del patrón

En los numerales 306 a 317 se presentan ejemplos de ocurrencias neutras según las áreas temáticas.

306. explicar trabajo (*general*)

307. generar encadenamiento (*general*)

308. llamar atención (*general*)

309. cambiar Constitución (*política*)

310. ofrecer Alcaldía (*política*)

311. realizar institución estatal (*política*)

- 312. mostrar índice de pobreza (*economía*)
- 313. medir desigualdad (*economía*)
- 314. pasar economía doméstica (*economía*)
  
- 315. destacar lumbalú (*sociedad*)
- 316. ver montaña (*sociedad*)
- 317. vivir mar (*sociedad*)

### 6.1.15. RECAPITULACIÓN

En este capítulo se ha presentado el análisis de las ocurrencias los patrones [verbo + objeto directo], recuperadas de un corpus sobre pobreza en la prensa colombiana, con propósitos de AS. A continuación, se presenta una síntesis de los hallazgos más relevantes, resultado de nuestro análisis.

1. En primer lugar, las 1.360 ocurrencias analizadas corresponden al 50 % de los datos extraídos, luego de las etapas de depuración y validación por medio de los *Scripts*. A este patrón [verbo + objeto directo] se les asignó la información correspondiente al tipo y sentido para los verbos; y el área temática para los objetos directos. Con relación a la polaridad, se definieron tres variables discretas: *positivo* (1), *negativo* (-1) y *neutro* (0). Se etiquetó cada elemento conforme a la carga correspondiente y al final, se definieron las posibles combinaciones según la estructura [verbo con carga X + objeto directo con carga Y = carga final patrón].
2. En segundo lugar, frente al tipo de verbo, los más frecuentes son los fraseológicos (60,67 %); seguidos de los verbos discursivos (24,72 %); por último, los verbos conectores (14,61 %). Según Lorente (2002), los verbos conectores ayudan en la construcción del discurso especializado; en consecuencia, los artículos analizados en este trabajo pertenecen a este tipo de discurso (la pobreza), dado que los verbos temáticos utilizados corresponden al 75,28 % del total.

3. En tercer lugar, sobre el sentido del verbo, estos se refieren al modo como se aborda la temática, puesto que permiten establecer un esquema alrededor de la descripción y evolución del tema en cuestión. En esta investigación, el 61,03 % de los datos dan cuenta de cuatro sentidos verbales. Los de *possession* son los más frecuentes (17,28 %); seguidos de los de *communication* (15,74 %); *change* (14,44 %) y *social* (12,57 %).
4. En cuarto lugar, con relación al área temática de los objetos directos, se ha identificado que la pobreza en Colombia sería abordada desde lo económico, lo social y lo político. Este orden se puede establecer gracias a la frecuencia del área temática: *economía* (29,19 %); *sociedad* (20,51 %) y *política* (19,04 %). Al respecto, las ocurrencias agrupadas en el área *general* (49,49 %) corresponden a estrategias discursivas.
5. En quinto lugar, con referencia a la carga de los verbos, el fenómeno de la pobreza es descrito mediante el uso predominante de verbos positivos, los cuales dan cuenta del 51,69 % de los datos; mientras que los negativos corresponden al 34,83 % y los neutros, al 13,48 %.
6. En sexto lugar, en cuanto a la carga de los objetos directos, estos son en su mayoría neutros (49,49 %). Como se ha establecido, lo neutro está más cerca de ser positivo que de ser negativo. En consecuencia, los complementos directos utilizados en la representación de la pobreza son positivos en un 80,45 %, teniendo en cuenta las 1.360 ocurrencias analizadas.
7. Por último, con los datos sobre la carga final del patrón [verbo + objeto directo], se podría establecer que el fenómeno de la pobreza se representa positivamente en la prensa colombiana. El 65,44 % de las ocurrencias son positivas, el 17,94 % negativas y el 16,62 % neutras. Teniendo en cuenta el comportamiento del neutro, la pobreza en Colombia se representa positivamente, lo que se evidencia con el 82,06 % de las ocurrencias analizadas (1.116 de 1.360).

De esta forma, se podría concretar que, por un lado, la pobreza en Colombia es un fenómeno representado positivamente en la prensa y, por otro lado, una adecuada clasificación de las ocurrencias del patrón [verbo + objeto directo] permitiría mejorar la clasificación de la polaridad de los textos por medio de la aplicación de herramientas de AS.

## **PARTE V: CONCLUSIONES**



## CAPÍTULO 7: CONCLUSIONES

### 7.1. CONCLUSIONES

En el siguiente capítulo se presentan las conclusiones, producto del análisis de los datos y las discusiones con relación a las hipótesis y a los objetivos planteados. Así mismo, se describen las dificultades enfrentadas durante el desarrollo de la investigación, los avances que esta representa y, en último lugar, las perspectivas de trabajo.

Esta investigación consistió en la clasificación de patrones [verbo + objeto directo] extraídos de artículos de prensa colombiana sobre pobreza, y la descripción de los hallazgos con miras a establecer la forma como es representado este tema en Colombia, a partir del análisis lingüístico de las ocurrencias de este patrón, unidas a elementos de otras áreas del conocimiento implicadas, como la economía, la sociología y la política, principalmente.

Aunque el presente trabajo no es una investigación sobre AS, los resultados encontrados están en directa relación con esta tarea, toda vez que permitirían una mejoría en la construcción de las herramientas para propósitos de AS en español. Además, los textos que hemos definido como objeto de estudio, han sido poco trabajados, en concordancia con la literatura consultada sobre el área, al ser escritos en español y pertenecer al campo de la socio-política. Por último, la unidad de análisis, el verbo más el objeto directo, también ha sido poco estudiada desde el AS, puesto que se han privilegiado otras categorías gramaticales, en especial, el adjetivo. Los hallazgos de esta investigación, permitieron ofrecer una definición de lo neutro, polaridad que por su naturaleza ha sido poco trabajada y rara vez definida por los investigadores desde los trabajos en esta área.

En primer lugar, se abordan los resultados con relación a los objetivos, para luego centrarnos en la descripción de las hipótesis a la luz de los hallazgos. En segundo lugar, se plantea el aporte de esta investigación al campo de estudio. Finalmente, se abordan las perspectivas de trabajo.

Las conclusiones aquí presentadas, más que generalizaciones sobre la temática analizada, tienen un carácter descriptivo en cuanto permiten, guardando los límites, establecer el comportamiento del fenómeno estudiado de manera interna. Por ende, no hay una validación externa de los datos.



### 7.1.1. VALIDACIÓN DE LOS OBJETIVOS

Con respecto al objetivo general de la investigación que es diseñar una metodología para la construcción de una base de datos con patrones [verbo + objeto directo] que permita la discriminación de estas como unidad lingüística para el mejoramiento de la clasificación automática de artículos de prensa por medio de las técnicas de análisis de sentimientos, consideramos que la guía metodológica propuesta es importante en el diseño de herramientas en español para esta tarea, teniendo en cuenta que permite una clasificación de los elementos discriminantes de la polaridad con relación a la opinión expresada en los artículos. Además, nuestra propuesta es susceptible de ser retomada desde cualquier área de trabajo donde se busque identificar la polaridad contenida en la opinión alrededor del tema trabajado.

En cuanto al primer objetivo específico que es caracterizar los patrones [verbo + objeto directo] para propósitos de Análisis de Sentimientos, identificando sus rasgos distintivos y discriminantes, hemos logrado identificar los elementos discriminantes de este patrón y clasificarlos de acuerdo a las características individuales.

Con relación al verbo, hemos logrado establecer una tipología conforme a los trabajos de Lorente (2002) sobre el uso de este en el discurso especializado, al igual que hemos logrado establecer el sentido de los verbos de nuestro corpus. Con relación a este aspecto, hemos hallado que las herramientas para análisis de sentimientos en español, en especial *SentiWordNet*, presenta problemas a la hora de la desambiguación de sentidos entre el inglés y el español, cuando el verbo es tomado como unidad de análisis. Esto nos ha permitido concluir que es necesaria la construcción de una base de datos para propósitos de AS propia del español; es decir, una base de datos donde el idioma principal sea el español y no se tenga como base otros idiomas como es la regularidad en el área, donde *WordNet*, cuyo idioma base es el inglés, se ha convertido en el recurso lexical predilecto para la construcción de herramientas en otros idiomas, entre ellos el español, mediante la técnica de mapeo. Sin embargo, estos hallazgos sobre *Wordnet*, como recurso lexical, no implican que este se constituya como un elemento indispensable en el campo, prueba de ello es su amplia utilización por parte de los investigadores en el área.

Unido a la clasificación del tipo y sentido del verbo, hemos logrado clasificar los verbos del corpus en las distintas polaridades trabajadas: positivo, negativo o neutro. Esta clasificación nos

ha permitido observar que, el verbo como unidad de estudio para propósitos de AS, al igual que el adjetivo, podría ser una categoría discriminante en polaridad y, además, junto a sus objetos directos, podría permitir la identificación de la tendencia de opinión del texto analizado y, contribuir con la clasificación final de los textos, de acuerdo a la carga de opinión expresada por el autor. Puesto que no hemos hecho experimentos corriendo programas de AS ni hemos realizado análisis estadísticos al respecto, no podemos extrapolar estas afirmaciones en un sentido más computacional.

Respecto a los objetos directos, hemos logrado establecer una clasificación de los mismos de acuerdo a sus áreas temáticas, lo que nos ha permitido concluir que, mediante su categorización, se podrían establecer los temas tratados en los artículos analizados, los cuales van en consonancia con las secciones de la prensa de donde fueron compilados. Además, hemos hallado que, con la asignación de la carga de polaridad a los objetos directos, se podría construir una tendencia en la representación del tema en la prensa colombiana, en términos de opinión.

Por último, la combinación de las polaridades asignadas al verbo y al objeto directo mediante la regla [verbo con carga X + objeto directo con carga Y = carga final del patrón], podría permitir el establecimiento de las tendencias con relación a la polaridad de la opinión expresada sobre el tema tratado en los artículos analizados, respecto al verbo como categoría discriminante. Ello constituiría un aporte importante en la construcción de una solución final a los desafíos propios del área del AS en categorías léxicas no típicas.

En cuanto al segundo objetivo específico que es definir una metodología de depuración de los patrones [verbo + objeto directo] de tal modo que puedan extraerse los elementos más discriminantes en la construcción de las herramientas para análisis de sentimientos, hemos logrado establecer una serie de criterios para la clasificación de estas ocurrencias y la asignación de las cargas de polaridad, criterios que consideramos, logran eliminar la información que no es relevante en términos de polaridad; y de este modo, lograr que en la construcción de bases de datos con el patrón [verbo + objeto directo], se tengan en cuenta solo los elementos discriminantes, depurando la información innecesaria que los sistemas interpretarían como ruido y podrían, de cierto modo, influir en los resultados del análisis.

Con relación a los elementos que hemos eliminado de nuestros datos de trabajo, hemos establecido algunas ideas al respecto. En primer lugar, en cuanto al ruido de la extracción de los patrones correspondientes a las ocurrencias que categorizamos como *errores de extracción, formas impersonales o formas incompletas*, hemos hallado que la eliminación de estas permite que, al momento en que se ejecuten los programas de análisis de sentimientos no sean tenidas en cuenta. Bajo los criterios de extracción y análisis, estas ocurrencias no aportan ningún tipo de información sobre el tema analizado.

En segundo lugar, con relación a las ocurrencias eliminadas categorizadas como *pronombres, deícticos y nombres propios*, hemos concluido que estos elementos no aportan información sobre la polaridad. Sin embargo, sí aportan información relevante para la construcción contextual de los artículos; es decir, por medio de estas se recupera la información correspondiente al sujeto de la oración al igual que información circunstancial sobre lugares, tiempo, entre otros.

En tercer lugar, con las ocurrencias repetidas hemos tomado la decisión de dejar una sola por grupo, teniendo en cuenta que para la caracterización basta con una sola para llevar a cabo el análisis. En cuarto lugar, los verbos modales no se tuvieron en cuenta en nuestro análisis, puesto que, por su naturaleza sintáctica (no son verbos semánticos), no recae sobre ellos expresar la polaridad, pero auxilian al verbo principal encargado de hacerlo. Además, al estar siempre acompañados de otro verbo, no cumplen los criterios de extracción del patrón [verbo + objeto directo] que hemos definido para nuestro análisis.

Por último, las ocurrencias que hacen referencia o que contienen cifras, las hemos excluido de nuestro análisis, puesto que consideramos deben ser tratadas en un trabajo de investigación futuro, debido al gran número de elementos que implican, además de los desafíos semánticos que estas ocurrencias representan y, por ende, la definición de las características para poder ser automatizadas, requiere un cuidado y dedicación especial.

Con relación al tercer y último objetivo específico que es describir cómo se aborda el tema de la pobreza en la prensa colombiana mediante la descripción lingüística de los patrones [verbo + objeto directo] para propósitos de Análisis de Sentimientos, se podría lograr, mediante el análisis de estos patrones, clasificados y etiquetados, establecer cómo se representa el fenómeno de la pobreza en la prensa colombiana en el periodo comprendido entre los años 2010 a 2014.

Teniendo presente lo que hemos establecido en cuanto al neutro y su tendencia positiva en los datos, se podría afirmar que la pobreza en Colombia es un fenómeno que, en el periodo mencionado, fue tratado positivamente. Hemos llegado a esta conclusión teniendo en cuenta los 1.360 patrones [verbo + objeto directo] analizados, de los cuales, 890 son positivos; 226 son neutros y 244 son negativos. Así mismo, de los 89 verbos analizados (véase Anexo A), el 34,83 % son positivos, mientras que el 13,48 % son negativos; es decir, 31 verbos positivos frente a 12 negativos. Sin embargo, es necesario ampliar la cantidad de ocurrencias de análisis y corroborar esta afirmación mediante la implementación de los datos en sistemas de AS.

Con respecto a los verbos neutros, hemos establecido dos elementos que nos han permitido plantear una definición de esta polaridad. En primera instancia, con base en el etiquetaje de los tipos de verbo, unido a la carga asignada a estos, hemos identificado que el neutro cumple con una función discursiva; es decir, no aporta información sobre la polaridad, pero sí contribuye en la construcción del texto. Los verbos etiquetados como discursivos fueron igualmente marcados con carga neutra. En segunda instancia, según las características de los verbos fraseológicos y conectores en Lorente (2002), donde estos no son portadores de lenguaje especializado sino que forman parte en la construcción del mismo, nos ha permitido definir el neutro con relación a la ausencia de contenido temático.

### **7.1.2. VALIDACIÓN DE LAS HIPÓTESIS**

Con relación a las hipótesis de nuestro trabajo y en consonancia con los objetivos que hemos descrito en el numeral anterior, presentamos las siguientes precisiones.

Con respecto a la primera hipótesis de nuestro trabajo, en la cual planteamos que el verbo, al igual que el adjetivo, es un elemento gramatical poseedor de opinión y, por lo tanto, debe ser tenido en cuenta a la hora de diseñar herramientas para Análisis de Sentimientos; esta se ha validado parcialmente, teniendo en cuenta que con las ocurrencias de los patrones [verbo + objeto directo] que hemos analizado, nos ha sido posible establecer tendencias con relación a la polaridad de los artículos. Sin embargo, de nuevo se hace necesario ampliar los datos de trabajo e implementarlos en un sistema de AS, con el fin de validar o no esta primera hipótesis de investigación.

Las ocurrencias del patrón [verbo + objeto directo] que hemos analizado, nos han permitido establecer una serie de criterios para la clasificación de los elementos más discriminantes y de esta manera, contribuir con la solución de los desafíos propios del área del AS, al presentar una guía metodológica para la caracterización de estos patrones sobre pobreza en Colombia.

Con respecto a la segunda hipótesis, en la cual planteamos que a partir del análisis lingüístico de las ocurrencias del patrón [verbo + objeto directo] es posible identificar la forma como se aborda el fenómeno de la pobreza en la prensa colombiana, esta no fue validada, toda vez que no fue posible implementar los datos en un sistema de AS y correr los análisis por medio de los cuales sería posible corroborar nuestra segunda hipótesis de trabajo.

## **7.2. APORTE DE LA INVESTIGACIÓN AL CAMPO DE ESTUDIO**

Los resultados de este trabajo de investigación son relevantes dado a que constituyen un aporte a la solución de la clasificación automática de los textos a partir del AS. Desde la descripción lingüística del verbo y su objeto directo, hemos logrado presentar una guía metodológica con este patrón como unidad de análisis, lo que nos ha permitido establecer que el verbo, como unidad lingüística, unido al objeto directo, es discriminante en términos de polaridad, al igual que otras categorías lingüísticas como el adjetivo; lo que amplía el panorama de trabajo a otras unidades de análisis poco estudiadas. Además, los textos de carácter socio-político que hemos analizado, han sido poco trabajados con fines de AS, lo que también constituye un aporte en la descripción de estos para tales propósitos.

De igual forma, nuestra propuesta (con los ajustes necesarios), puede ser aplicada o extrapolada a otros fenómenos que se deseen investigar. Por último, a partir de los datos analizados, hemos presentado un acercamiento teórico al neutro como variable de polaridad, presentando una definición con base en los resultados obtenidos, lo que constituye un aporte al área del AS, en la construcción de una tipología del neutro, variable que en la literatura no ha sido abordada ampliamente, consecuencia de los desafíos que implica una clasificación triclase (positivo, negativo, neutro) de los textos.

### **7.3. LIMITACIONES Y PERSPECTIVAS DE TRABAJO**

Todo trabajo de investigación presenta limitaciones. En este caso, la principal limitación tiene que ver con la amplitud de la muestra analizada y de allí que los patrones [verbo + objeto directo] sean limitados. En consecuencia, el análisis y los resultados presentados no pueden ser generalizables y son más bien predictivos del fenómeno que hemos analizado.

Una segunda limitación de nuestro trabajo, tiene que ver con la poca información en la literatura con relación a nuestra unidad de estudio para propósitos de AS, lo que no ha permitido comparar estos resultados con otros trabajos y así, poder establecer conclusiones o generalizaciones. Sin embargo, esta limitación también puede ser considerada como un aporte de la investigación, en cuanto abre el camino para que futuras investigaciones retomen nuestra unidad de análisis y de esta forma, se puedan extrapolar los resultados aquí presentados.

Este trabajo de investigación, nos ha permitido establecer varias líneas de trabajo, de las cuales mencionaremos algunas. Para que nuestros resultados sean generalizables, es necesario llevar a cabo una investigación donde pueda ser analizado un corpus más amplio y observar, de esta manera, si el patrón [verbo + objeto directo] es lo suficientemente discriminante en las aplicaciones de AS, desde el punto de vista lingüístico y computacional.

Teniendo en cuenta la revisión bibliográfica que hemos realizado junto a los resultados del análisis de nuestros datos, hemos identificado un vacío con relación a las herramientas para AS en español que, aunque existentes, no son lo suficientemente potentes para llevar a cabo la clasificación automática de la polaridad de los textos en esta lengua. De allí, la posibilidad de emprender un proyecto centrado en la construcción de una herramienta propia del español para propósitos de Análisis de Sentimientos.

## BIBLIOGRAFÍA

- Alarcos, E. (1999). *Gramática de la lengua española*. Madrid, España: Espasa.
- Alcoba, S. (1999). La flexión verbal. En Bosque, I., & Demonte, V. (Eds.). *Gramática Descriptiva de la Lengua Española* (p. 4917)
- Alkire, S., & Foster, J. (2008). Counting and Multidimensional Poverty Measurement. *Journal of Public Economics*, 95(7), 476-487.
- Alonso, L., Capilla, J. A., Castellón, I., Fernández-Montraveta, A., & Vázquez, G. (2005). The SenSem Project: Syntactico-Semantic Annotation of Sentences in Spanish. In Nicolov, N., Bontcheva, K., Angelova, G., & Mitkov, R. (Eds.). *Recent Advances in Natural Language Processing IV* (pp. 89–98). Amsterdam, The Netherlands: John Benjamins Publishing.
- Álvarez, S., & Naharro, N. (2017). Representaciones de la pobreza en la prensa hegemónica Argentina. En Chiquito, A., & Quiroz, G. (Eds.). *Pobreza, Lenguaje y Medios en América Latina* (pp. 235–264). Berna, Suiza: Peter Lang.
- Arnauld, A. (1846). *Grammaire générale et raisonnée de Port- Royal*. Paris, France: Imprimerie de Munier.
- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. New York, USA: ACM Press.
- Bagheri, A., Saraee, M., & de Jong, F. (2013). Latent Dirichlet Markov Allocation for Sentiment Analysis. *The Fifth European Conference on Intelligent Management Systems in Operations (IMSIO 5)*, 90-96.
- Bello, A. (1887:2014). *Gramática de la Lengua Castellana: destinada al uso de los americanos*. Barcelona, España: Linkgua Digital.
- Benamara, F., Cesarano, C., Picariello, A., Reforgiato, D., & Subrahmanian, V. (2007). Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone. *In Proceedings of International Conference on Weblogs and Social Media*.
- Benveniste, É. (1966:1977). *Problemas de lingüística general I*. Coyoacán, México: Siglo Veintiuno Editores.
- Blei, D. M., Edu, B. B., Ng, A. Y., Edu, A. S., Jordan, M. I., & Edu, J. B. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, (3), 993-1022.

- Blei, D. M., & McAuliffe, J. D. (2010). Supervised Topic Models. In Lafferty, J.D., Williams, C.K.I., Shawe-Taylor, J., Zemel, R.S., & Culotta, A. (Eds.). *Advances in Neural Information Processing Systems* (pp. 121-128). New York, USA: Curran Associates, Inc.
- Bloomfield, L. (1983). *An Introduction to the Study of Language* (3). New York, USA: Henry Holt and Company.
- Bosque, I. (2006). *Diccionario Combinatorio Práctico del Español Contemporáneo*. Madrid, España: Ediciones SM.
- Cabral, L., & Hortaçsu, A. (2004). The Dynamics of Seller Reputation: Evidence from eBay. *The Journal of Industrial Economics*, 58(1), 54-78.
- Carvajal, H. (1992). El concepto de verbo en la tradición gramática española. *Logos: Revista de Lingüística, Filosofía y Literatura* (6-7), 1-20.
- Castro, L. T. de, & Rodríguez, J. (1974). Algunas consideraciones sobre el aspecto verbal en español. *Thesaurus: Boletín del Instituto Caro y Cuervo*, 39(1), 34-47.
- Chareaudeau, P. (1997). *Le discours d'information médiatique. La construction du miroir social*. Paris, France: Institut national de l'audiovisuel.
- Chiquito, A., & Quiroz, G. (Eds.) (2017). *Pobreza, Lenguaje y Medios en América Latina*. Berna, Suiza: Peter Lang.
- Chomsky, N. (1997). *What Makes Mainstream Media Mainstream*. Retrieved December 20, 2017, from <http://www.zmag.org/chomsky/articles/z9710-mainstream-media.html>
- Crego, M. (2003). Algunas observaciones en torno a la valencia verbal. *Revista de Filología de La Universidad de La Laguna*, (21), 69-89.
- Cruz, F., Troyano, J., Pontes, B., & Ortega, F. J. (2014). ML-SentiCon: Un Lexicón Multilingüe de Polaridades Semánticas a Nivel de Lemas. *Procesamiento Del Lenguaje Natural*, 53, 113-120.
- Colombia. DANE. (2012). *Pobreza en Colombia*. Bogotá D.C., Colombia: Oficina de Prensa DANE.
- Dapena, J. Á. P. (2002). *Complementos argumentales del verbo: directo, indirecto, suplemento y agente*. Madrid, España: Arco Libros.
- Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*, 53(9), 1375-1388.



- Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In *Proceedings of the 12th International Conference on World Wide Web* (pp. 519-528). ACM.
- De Nebrija, A. (1492:2010). *Gramática de la lengua castellana*. Barcelona, España: Linkgua Digital.
- De Saussure, F., Bailly, C., & Séchehaye, A. (1919:1967). *Cours de linguistique générale. Collection Payothèque*. Paris, France: Payot & Rivages.
- Deng, L. (2015). MPQA 3.0: An Entity/Event-Level Sentiment Corpus. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 2015* (pp. 1323-1328).
- Ding, X., Liu, B., & Yu, P. S. (2008). A Holistic Lexicon-based Approach to Opinion Mining. In *Proceedings of the International Conference on Web Search and Web Data Mining - WSDM '08* (pp. 231-240).
- Dong, Z., & Dong, Q. (2003). HowNet - A Hybrid Language and Knowledge Resource. In *NLP-KE 2003 - 2003 International Conference on Natural Language Processing and Knowledge Engineering, Proceedings* (pp. 820-824).
- Fillmore, C. J. (1967). The Case for Case. In *Proceedings of the Texas Symposium on Linguistics Universals* (pp. 1-134).
- Gamon, M. (2004). Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors, and the Role of Linguistic Analysis. In *Proceedings of the 20th International Conference on Computational Linguistics* (p. 841). Association for Computational Linguistics.
- García-Miguel, J. M. (2012). Lingüística de corpus y valencia verbal. En Moskowich, I., & Crespo, B. (Eds.). *Encoding the Past, Decoding the Future: Corpora in the 21st Century* (pp. 29-57). Newcastle, UK: Cambridge Scholars.
- García, M. D. (1996). El verbo en algunos gramáticos latinos: sus accidentes. *Cuadernos de Filología Clásica. Estudios Latinos*, (10), 45-63.
- Gili Gaya, S. (1980). *Curso superior de sintaxis española*. Barcelona, España: Vox Bibliograf.
- González Calvo, J. M. (1978). El concepto de verbo. *Anuario de Estudios Filológicos*, 65-90.
- Grijelmo, A. (2006). *La Gramatica Descomplicada*. Ciudad de México, México: Taurus.
- Hagedorn, B. A., Ciaramita, M., & Atserias, J. (2007). World knowledge in Broad-coverage Information Filtering. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 801-802). ACM.

- Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the Semantic Orientation of Adjectives. *In Proceedings of the 35th Annual Meeting on Association for Computational Linguistics* (pp. 174–181).
- Hatzivassiloglou, V., & Wiebe, J. M. (2000). Effects of Adjective Orientation and Gradability on Sentence Subjectivity. *In Proceedings of the 18th Conference on Computational Linguistics*, (1), (pp. 299-305). Association for Computational Linguistics.
- Hearst, M. A. (1992). Direction-Based Text Interpretation as an Information Access Refinement. In Jacobs, P. (Ed.). *Text-Based Intelligent Systems* (pp. 257-274). New York, USA: Psychology Press.
- Helbig, G. (1992). *Probleme der Valenz- und Kasustheorie*. Tübingen, Deutschland: Max Niemeyer Verlag.
- Helena, L. (2016). La semántica. En Gutiérrez-Rexach, J. (Ed.). *Enciclopedia de lingüística hispánica* (pp. 284-300). Abingdon, Reino Unido: Routledge.
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación*. Ciudad de México, México: McGraw-Hill.
- Hoffmann, P., Wilson, T., & Wiebe, J. (2009). Recognizing Contextual Polarity: An Exploration of Features for Phrase-level Sentiment Analysis. *Computational Linguistics*, 35(3), 399-433.
- Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews. *In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 168-177). ACM.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- Kennedy, A., & Inkpen, D. (2006). Sentiment Classification of Movie and Product Reviews Using Contextual Valence Shifters. In Engelbrecht, A. P. (Ed.). *Computational Intelligence: an Introduction* (pp. 110–125). Chichester, England: John Wiley & Sons.
- Kim, S.-M., & Hovy, E. (2006). Extracting opinions, Opinion Holders, and Topics Expressed in Online News Media Text. *In Proceedings of the Workshop on Sentiment and Subjectivity in Text* (pp. 1-8). Association for Computational Linguistics.
- Kim, S.-M., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically Assessing Review Helpfulness. *In Proceedings of the 2006 Conference on empirical methods in natural language processing* (pp. 423-430). Association for Computational Linguistics.

- Laca, B. (2006). El objeto directo. La marcación preposicional. In Company, C. (Ed.) *Sintaxis histórica de la Lengua Española. Primera Parte: la frase verbal* (pp. 423-478). Ciudad de México, México: Fondo de Cultura Económica.
- Levin, B. (1993). *English Verb Classes and Alternations*. Chicago, USA: The University of Chicago Press.
- Li, F., Huang, M., & Zhu, X. (2007). Sentiment Analysis with Global Topics and Local Dependency. In *Association for the Advancement of Artificial Intelligence* (10) (pp. 1371-1376).
- Liang, J., Liu, P., Tan, J., & Bai, S. (2014). Sentiment Classification Based on AS-LDA Model. *Procedia Computer Science*, (31), 511-516.
- Liu, B. (2010). Sentiment Analysis and Subjectivity. In Dale, R., Moisl, H., & Somers, H. (Eds.). *Handbook of Natural Language Processing* (pp. 627-666). Cambridge, UK: CRC Press.
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. In Hirst, G. (Ed.). *Synthesis Lectures on Human Language Technologies* (pp. 1-167). California, USA: Morgan & Claypool.
- Liu, B. (2015). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. *Computational Linguistics* (42). New York, USA: Cambridge University Press.
- Lorente, M. (2002). Verbos y discurso especializado. *Estudios de Lingüística del Español* [En línea]. Recuperado de <http://elies.rediris.es/elies16/Lorente.html>
- M. Airoidi, E., Bai, X., Padman, R., & Airoidi, E. (2006). Markov Blankets and Meta-heuristics Search: Sentiment Extraction from Unstructured Texts. In *International Workshop on Knowledge Discovery on the Web* (pp. 167-187). Berlin, Heidelberg: Springer.
- Marzá, N. E. (2012). Lexicografía especializada y lenguajes de especialidad: Fundamentos teóricos y metodológicos para la elaboración de diccionarios especializados. *Lingüística*, 27(1), 98-135.
- Matsumoto, S., Takamura, H., & Okumura, M. (2005). Sentiment Classification Using Word Subsequences and Dependency Sub-trees. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (5), (pp. 301-311).
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. New York, USA: Cambridge University Press.
- McEnery, T., & Wilson, A. (2001). *Corpus Linguistics: An Introduction*. Edimburgh, Scotland: Edinburgh University Press.
- Mcluhan, M., & Fiore, Q. (1988). *El medio es el masaje. Un inventario de efectos*. Barcelona, España: Paidós Ibérica.

- Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3(4), 235-244.
- Moliner, M. (2007). *Diccionario de uso del español*. Madrid, España: Gredos.
- Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. (2002). Mining Product Reputations on the Web. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 341-349). ACM.
- Mullen, T., & Collier, N. (2004). Sentiment Analysis Using Support Vector Machines with Diverse Information Sources. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (4), (pp. 412-418).
- Na, J., Sui, H., Khoo, C., Chan, S., & Zhou, Y. (2004). Effectiveness of Simple Linguistic Processing in Automatic Sentiment Classification of Product Reviews. In *Advances in Knowledge Organization* (9), (pp. 49-54).
- Naciones Unidas. (2000). Declaración del Milenio. New York, (55/2), 1-17. Recuperado de [http://www2.ohchr.org/spanish/bodies/hrcouncil/docs/gaA.RES.60.1\\_Sp.pdf](http://www2.ohchr.org/spanish/bodies/hrcouncil/docs/gaA.RES.60.1_Sp.pdf)
- Naciones Unidas. (2015). Objetivos de Desarrollo del Milenio. Nuevas Ediciones S.A. Recuperado de [http://www.un.org/es/millenniumgoals/pdf/PR\\_Global\\_MDG09\\_SP.pdf](http://www.un.org/es/millenniumgoals/pdf/PR_Global_MDG09_SP.pdf)
- Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing Favorability Using Natural Language Processing. In *Proceedings of the 2nd International Conference on Knowledge Capture* (pp. 70-77). ACM.
- Ng, V., Dasgupta, S., & Arifin, S. M. N. (2006). Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews. In *Proceedings of the COLING/ACL on Main Conference Poster Sessions* (pp. 611-618). Association for Computational Linguistics.
- Oronich, O. M. (2009). La triple caracterización de la “oración” de Samuel Gili Gaya: una teoría vigente. *Revista de filología de la Universidad de La Laguna* (27), 109-122.
- Øyen, E. (2004). Poverty Production: a Different Approach to Poverty Understanding. In Genov, N. (Ed.). *Advances in Sociological Knowledge* (pp. 299-315). Heidelberg, Germany: Springer Science & Business Media.
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*. Hanover, USA: Now Publishers Inc.

- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, (10), (pp. 79-86). Association for Computational Linguistics.
- Pardo, N., & Ruiz, J. (2017). Pobreza y bienestar en Colombia, construcción de referenciales en perspectiva mediática. En Chiquito, A., & Quiroz, G. (Eds.). *Pobreza, Lenguaje y Medios en América Latina* (pp. 265-302). Berna, Suiza: Peter Lang.
- Picard, R. W. (2003). Affective Computing: Challenges. *International Journal of Human-Computer Studies*, 59(1), 55-64.
- Quiroz, G. (2003). Preparación y procesamiento de un corpus para la creación de materiales en clase de español para propósitos específicos. En *Actas del X Congreso Brasileño de Profesores de Español* (pp. 131-150). Natal, Brasil: Thesaurus Editora de Brasilia.
- Quiroz, G., Chiquito, A., & Zuluaga, J. (2017). Cómo se representa lingüísticamente la pobreza en la prensa colombiana. En Chiquito, A. & Quiroz, G. (Eds.). *Pobreza, Lenguaje y Medios en América Latina* (pp. 83-107). Berna, Suiza: Peter Lang.
- Real Academia Española. (1771). Gramática de la Lengua Castellana. Recuperado de [http://www.rae.es/sites/default/files/Gramatica\\_RAE\\_1771\\_reducida.pdf](http://www.rae.es/sites/default/files/Gramatica_RAE_1771_reducida.pdf)
- Real Academia Española. (2010). *Nueva Gramática de la Lengua Española*. Barcelona, España: Espasa.
- Real Academia Española. (2017). *Diccionario de la Lengua Española*. Madrid, España: Espasa.
- Riloff, E., Patwardhan, S., & Wiebe, J. (2006). Feature Subsumption for Opinion Analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (pp. 440-448). Association for Computational Linguistics.
- Riloff, E., & Wiebe, J. (2003). Learning Extraction Patterns for Subjective Expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing* (pp. 105-112). Association for Computational Linguistics.
- Riloff, E., Wiebe, J., & Wilson, T. (2003). Learning Subjective Nouns Using Extraction Pattern Bootstrapping. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003* (4) (pp. 25-32). Association for Computational Linguistics.
- Robins, R. H. (1997). *Breve historia de la lingüística*. Madrid, España: Ediciones Cátedra.

- Schulte im Walde, S. (2008). The Induction of Verb Frames and Verb Classes from Corpora. In Lüdeling, A., & Kytö, M. (Eds.). *Corpus Linguistics: an International Handbook* (2nd Ed.) (pp. 952-971). Berlin, Germany: Walter de Gruyter.
- Snyder, B., & Barzilay, R. (2007). Multiple Aspect Ranking Using the Good Grief Algorithm. In *Proceedings of NAACL HLT* (pp. 300-307).
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based Methods for Sentiment Analysis. *Computational linguistics*, 37(2), 267-307.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris, France: Librairie C. Klincksieck.
- Tokuhisa, R., Inui, K., & Matsumoto, Y. (2008). Emotion Classification Using Massive Examples Extracted from the Web. In *Proceedings of the 22nd International Conference on Computational Linguistics*, (1), (pp. 881-888). Association for Computational Linguistics.
- Turney, P. D. (2002). Thumbs up or Thumbs down?: Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (pp. 417-424). Association for Computational Linguistics.
- Vendler, Z. (1967). *Linguistic in Philosophy*. New York, USA: Cornell University Press.
- Villalobos, M. E. (1991). Las voces verbales del español. *Letras*, 1(25-26), 221-234.
- Wang, C., Blei, D., & Fei-Fei, L. (2009). Simultaneous Image Classification and Annotation. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 1903-1910).
- Whitelaw, C., Garg, N., & Argamon, S. (2005). Using Appraisal Groups for Sentiment Analysis. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management* (pp. 625-631). ACM.
- Wiebe, J. (1994). Tracking Point of View in Narrative. *Computational Linguistics*, 20(2), 233-287.
- Wiebe, J. M. (1990). Identifying Subjective Characters in Narrative. In *Proceedings of the 13th Conference on Computational Linguistics* (2), (pp. 401-406). Association for Computational Linguistics.
- Wiebe, J. M. (2000). Learning Subjective Adjectives from Corpora. In *Proceedings of the National Conference on Artificial Intelligence* (pp. 735-741).
- Wiebe, J. M., Bruce, R. F., & O'Hara, T. P. (1999). Development and Use of a Gold-standard Data Set for Subjectivity Classifications. In *Proceedings of the 37th Annual Meeting of the Association for*

*Computational Linguistics on Computational Linguistics* (pp. 246-253). Association for Computational Linguistics.

Wüster, E. (2010). *Introducción a la teoría general de la terminología y lexicografía terminológica*. Barcelona, España: Documenta Universitaria.

Yin, S., Han, J., Huang, Y., & Kumar, K. (2014). Dependency-topic-affects-sentiment-LDA Model for Sentiment Analysis. In *Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on Tools with Artificial Intelligence* (pp. 413-418). IEEE.

Zamorano, A. (2002). El verbo en las gramáticas del español (de Andrés Bello a Samuel Gili Gaya). Definición y clasificaciones tipológicas. *Anuario de Letras: Lingüística y filología*, 40, 205-240.

## ANEXOS

### Anexo A: Frecuencia de verbos en la muestra

Frecuencia verbos		
Verbo	Frecuencia	%
hacer	78	5,74
dar	65	4,78
mejorar	46	3,38
generar	34	2,5
reducir	27	1,99
decir	26	1,91
explicar	26	1,91
recibir	25	1,84
ver	24	1,76
destacar	23	1,69
mostrar	22	1,62
poner	21	1,54
presentar	21	1,54
aumentar	19	1,4
hablar	19	1,4
superar	19	1,4
buscar	18	1,32
crear	18	1,32
mantener	18	1,32
realizar	18	1,32
tomar	18	1,32
afectar	16	1,18
contar	16	1,18
existir	16	1,18
indicar	16	1,18
necesitar	16	1,18
seguir	16	1,18
apoyar	15	1,1
ir	15	1,1
lograr	15	1,1
ofrecer	15	1,1
venir	15	1,1
vivir	15	1,1
brindar	14	1,03
cumplir	14	1,03
defender	14	1,03



dejar	14	1,03
disminuir	14	1,03
encontrar	14	1,03
entregar	14	1,03
garantizar	14	1,03
sacar	14	1,03
bajar	13	0,96
considerar	13	0,96
construir	13	0,96
desarrollar	13	0,96
evitar	13	0,96
fortalecer	13	0,96
llevar	13	0,96
pasar	13	0,96
perder	13	0,96
sufrir	13	0,96
creer	12	0,88
llegar	12	0,88
pagar	12	0,88
salir	12	0,88
asegurar	11	0,81
cerrar	10	0,74
enfrentar	10	0,74
impulsar	10	0,74
incluir	10	0,74
llamar	10	0,74
obtener	10	0,74
resaltar	10	0,74
trabajar	10	0,74
cambiar	9	0,66
comprar	9	0,66
facilitar	9	0,66
incrementar	9	0,66
medir	9	0,66
permitir	9	0,66
producir	9	0,66
reconocer	9	0,66
señalar	9	0,66
acceder	8	0,59
crecer	8	0,59
establecer	8	0,59
ganar	8	0,59
implementar	8	0,59

integrar	8	0,59
pedir	8	0,59
preocupar	8	0,59
quedar	8	0,59
responder	8	0,59
saber	8	0,59
aprobar	7	0,51
asumir	7	0,51
combatir	7	0,51
avanzar	4	0,29
Palabras o elementos distintos contados: 89		
Total:	1.360	100

**Anexo B: Reglas de combinación según las cargas positivas,  
negativas o neutras**

<b>Carga verbo</b>	<b>Carga objeto directo</b>	<b>Carga patrón</b>
POS	POS	POS
POS	POS	NEG
POS	POS	NEU
POS	NEG	POS
POS	NEG	NEG
POS	NEG	NEU
POS	NEU	POS
POS	NEU	NEG
POS	NEU	NEU
NEG	POS	POS
NEG	POS	NEG
NEG	POS	NEU
NEG	NEG	POS
NEG	NEG	NEG
NEG	NEG	NEU
NEG	NEU	POS
NEG	NEU	NEG
NEG	NEU	NEU
NEU	POS	POS
NEU	POS	NEG
NEU	POS	NEU
NEU	NEG	POS
NEU	NEG	NEG
NEU	NEG	NEU
NEU	NEU	POS
NEU	NEU	NEG
NEU	NEU	NEU

## Anexo C: Reglas de combinación en los datos con ejemplos

Reglas de combinación	Ejemplos
POS+POS=POS	<ul style="list-style-type: none"> <li>❖ apoyar paz</li> <li>❖ garantizar educación de alta calidad</li> <li>❖ producir resultado positivo</li> </ul>
POS+POS=NEG	<ul style="list-style-type: none"> <li>❖ (no) brindar gran beneficio</li> <li>❖ (no) dar subsidio</li> <li>❖ (no) poner educación</li> </ul>
POS+NEG=POS	<ul style="list-style-type: none"> <li>❖ (no) dar limosna</li> <li>❖ superar conflicto interno</li> <li>❖ trabajar población vulnerable</li> </ul>
POS+NEG=NEG	<ul style="list-style-type: none"> <li>❖ apoyar minería informal</li> <li>❖ facilitar abuso</li> <li>❖ producir tristeza</li> </ul>
POS+NEG=NEU	<ul style="list-style-type: none"> <li>❖ dar pesar</li> </ul>
POS+NEU=POS	<ul style="list-style-type: none"> <li>❖ apoyar metro de Bogotá</li> <li>❖ cumplir meta</li> <li>❖ mejorar condición de vida</li> </ul>
POS+NEU=NEG	<ul style="list-style-type: none"> <li>❖ (no) cumplir requisito</li> <li>❖ generar distribución</li> <li>❖ incrementar costo</li> </ul>
POS+NEU=NEU	<ul style="list-style-type: none"> <li>❖ construir futuro</li> <li>❖ ofrecer información</li> <li>❖ vivir conferencia</li> </ul>
NEG+POS=POS	<ul style="list-style-type: none"> <li>❖ dejar ayuda</li> <li>❖ dejar razón de esperanza</li> <li>❖ (no) perder subsidio</li> </ul>
NEG+POS=NEG	<ul style="list-style-type: none"> <li>❖ afectar empleabilidad</li> <li>❖ necesitar subsidio</li> <li>❖ reducir espacio democrático</li> </ul>
NEG+POS=NEU	<ul style="list-style-type: none"> <li>❖ cerrar venta de activo</li> <li>❖ dejar Unidad Permanente de Derechos Humanos</li> </ul>
NEG+NEG=POS	<ul style="list-style-type: none"> <li>❖ bajar línea de pobreza</li> <li>❖ dejar pobreza</li> <li>❖ evitar abandono</li> </ul>
NEG+NEG=NEG	<ul style="list-style-type: none"> <li>❖ afectar embarazo adolescente</li> <li>❖ (no) reducir la pobreza</li> <li>❖ sufrir escasez de alimento</li> </ul>
NEG+NEG=NEU	<ul style="list-style-type: none"> <li>❖ preocupar pobreza</li> <li>❖ preocupar vulneración</li> <li>❖ preocupar desplazamiento</li> </ul>
NEG+NEU=POS	<ul style="list-style-type: none"> <li>❖ bajar precio medicamento</li> <li>❖ evitar sonda</li> <li>❖ reducir costo</li> </ul>

NEG+NEU=NEG	<ul style="list-style-type: none"> <li>❖ evitar importación</li> <li>❖ necesitar cuidado</li> <li>❖ perder participación</li> </ul>
NEG+NEU=NEU	<ul style="list-style-type: none"> <li>❖ bajar tendencia</li> <li>❖ dejar tierra</li> <li>❖ preocupar encuesta</li> </ul>
NEU+POS=POS	<ul style="list-style-type: none"> <li>❖ buscar trabajo</li> <li>❖ hacer inversión</li> <li>❖ tomar control</li> </ul>
NEU+POS=NEG	<ul style="list-style-type: none"> <li>❖ (no) contar servicio de gas</li> <li>❖ (no) encontrar solución</li> <li>❖ realizar acuerdo</li> </ul>
NEU+POS=NEU	<ul style="list-style-type: none"> <li>❖ considerar regalo</li> <li>❖ explicar trabajo</li> <li>❖ señalar política de vivienda</li> </ul>
NEU+NEG=POS	<ul style="list-style-type: none"> <li>❖ encontrar descalabro</li> <li>❖ responder ataque</li> <li>❖ sacar miseria</li> </ul>
NEU+NEG=NEG	<ul style="list-style-type: none"> <li>❖ encontrar mayor deserción</li> <li>❖ llegar austeridad</li> <li>❖ tomar terrible decisión</li> </ul>
NEU+NEG=NEU	<ul style="list-style-type: none"> <li>❖ considerar evolución de la pobreza</li> <li>❖ hablar pobreza</li> <li>❖ medir pobreza</li> </ul>
NEU+NEU=POS	<ul style="list-style-type: none"> <li>❖ encontrar eco</li> <li>❖ hacer pronunciamiento</li> <li>❖ realizar investigación</li> </ul>
NEU+NEU=NEG	<ul style="list-style-type: none"> <li>❖ (no) pasar prueba</li> <li>❖ quedar camino</li> <li>❖ (no) tomar decisión</li> </ul>
NEU+NEU=NEU	<ul style="list-style-type: none"> <li>❖ buscar formula</li> <li>❖ hacer acta</li> <li>❖ ver montaña</li> </ul>