Scientific Research Publishing

Open Journal of Social Sciences, 2021, 9, 141-166
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Corpus Linguistics Representations on Age Groups in Light of Google Books

Bárbara Silva

São Paulo Technological State College, São Paulo, Brazil
Email: barbara.silva55@fatec.sp.gov.br

How to cite this paper: Silva, B. (2021). *Corpus* Linguistics Representations on Age Groups in Light of Google Books. *Open Journal of Social Sciences, 9,* 141-166.
https://doi.org/10.4236/jss.2021.98011

Received: May 15, 2021
Accepted: August 7, 2021
Published: August 10, 2021

Abstract

The goal of the research lays on identifying social representations around words associated with human beings in Google Books BYU Corpus, in a period of 208 years, from 1800 to 2008. In this paper, the main data findings of a corpus-based investigation are focused on the adjectives preceding such words as man, woman, adolescent, boy, girl, child, and teenager in database. By verifying patterns of association between each of these words and immediate collocates, it is possible to infer how these concepts are represented over time. First, queries were conducted in the Corpus. Second, adjectives were selected. Third, these were classified into semantic categories. Fourth, collocates were classified through sentiment analysis. Finally, major representations were inferred based on semantic categories and sentiment analysis scores. The word "children" showed different representations: medical, consisting of collocates such as asthmatic, disabled, religious and evaluative. We have concluded that over time, representations of age, health and race increased, while representations of innocence decreased. It can be applied that the collocates that appeared in the latter half-century compared to first half-century give an indication of the current representations. Finally, for children, these include hyperactive and disadvantaged, indicating a shift toward a "problematic" representation of children.

Keywords

Corpus Linguistics, Google Books, Human Beings, Social Representations, Human Identifications

1. Introduction

According to Stubbs (1996: p. 158), the use patterns of lexical items can signal the representation that these items assume in society: recurrent ways of speaking do not determine the thought. They offer representations conventional or family members of people and events through the filter and the crystallization of ideas besides providing prefabricated meanings through which they can be easily captured and broadcast. (Author's translation) 1 The study presented in Stubbs (1996) analyzed some of the main British cultural manifestations around words like English, Scottish and British. The author sought the most frequent placement of these words, that is, linguistic patterns formed by the presence of two words next to each other (usually separated by up to four other words), such as "British Empire". The author used a subset of the Cobuild corpus as the source of the data analyzed, with 130 million words, of which practically 70% come from journalistic records, books (fiction and non-fiction) and conversation in English. Stubbs (1996) shows how lexical co-occurrences identified in corpora can be indications or marks of representations, thus motivating the present search. There are other studies in Corpus Linguistics about representations that also establish the same relationship between the frequent use of certain linguistic patterns and the presence of representations, such as Baker (2014), Baker & Potts (2013) and Baker and Ellece (2011). Although there are already corpora-based studies of representations, there is no research that investigates representations around lexical items related to human being. Since there is no precedent for studies within the Linguistics of Corpus dedicated to the investigation of social representation in relation to human beings, the research reported here aims to fill this gap.

2. Background

In 2014, I observed children and adolescents in the Education and Health at the Federal University of the State of São Paulo undergoing a Professional Update Specialization Course for non-doctors, in order to understand what, from the point of view of this age group, these life stages (child, youth, teenager, elderly, etc.) could represent for patients. Then, already enrolled in the Postgraduate Program in Applied Linguistics and Language Studies (LAEL), under Professor Tony Berber Sardinha Coordination, I found in *Corpus* Linguistics the basis for investigating the different social representations of the human being time. This article reports on a study whose objective is to deal with the investigation of linguistic characteristics of human identification words in English, more specifically man (man), woman (woman), adolescent (teenager), adolescence, adult (adult), boy (boy), girl (girl), child (child), elderly (elderly), kid (child), teen (teenager) and teenager (teenager), as well as its plural forms, from data from Google Books that cover the 1800 to 2008. In addition to identifying the patterns of these words, the present study aimed at check for changes in these patterns over the twenty-one decades of the study. After having taken the decision to research such an important subject, I have realized how incredible it is to understand that human beings are not defined by their ages. They are actually only represented by their ages because in our current society (and it has been like this for the past 208 years) uses such a classification to determine how times pass by. There is a new term mentioned among distinct medias nowadays, which is someone who is "ageless", it is to say, someone that even though may be on their twenties, thirties or forties is someone not represented by their ages. It may bring some issues for discussion, because society tends to positively value more the people who seem to be younger, whether it is physically speaking or in their appearance and way to express themselves. Such valuations are shown in the study according to social representations seems in each age group.

Research Questions

When I started defining the research questions, the way I found the most suitable to clarify the sections to the final reader, lays on answering the following questions in the present research:

1) What representations can be identified in relation to the terms surveyed?

2) Is there a difference between the representations of the masculine and feminine terms? In addition, among children, teenagers and adults?

3) Is there a difference between the terms in relation to valuation (positive charge and negative)?

The main reason why only three questions have been considered for the research is mainly due to the fact that terms of gender throughout the study have proven to be a massive field of study for further scenarios. Also, when it comes to understanding the different stages among young childhood and young adult age groups, it is also a main subject of study that deserves to be studied in a deeper level.

In the following section, methodology it will be further explained and in the chapter of results it will also be answered according to the study.

3. Methods

This section is dedicated to explaining the method of the present research according to research questions mentioned above. In this study, we performed the analysis of the language patterns of a set of words, in English, referring to the human being and, from these patterns; we try to verify the representations associated with the different ways of referring to being a human being over time. A pattern can be identified if a combination of words occurs relatively frequently and if there is an associated meaning. The data used in this research deal with listings of occurrences of bigrams found in English-language publications indexed by Google Books. Google Books is a collection of millions of digitized publications by the Google Company from library collections around the world. The format of this database, as well as its extension, will be explained in the section about procedures. Bigrams, on the other hand, are sequences of two words placed side by side in a text. For example: "Brazilian poet", "young women" and "American men". Bigrams are available on the Google Books Ngrams website, which also allows the user to search and produce graphs of the use of these bigrams. Thus, strictly speaking, not directly dealt with a corpus of texts, because the texts of publications indexed by Google Books are not made available to users. Google Books only offers bigram.

Therefore, our data were bigrams, along with their frequency of occurrence in publications in English indexed between 1800 and 2008. The following tools were used in this research, which are described in procedures section:

1) The Brigham Young University interface for Google Books N-Grams and this interface helps with searches on Google Books N-Grams.

2) Google Book's viewer N-Gram Viewer. Such online interface allows graphical display of N-gram occurrences in the database Google Books N-Gram data.

3) USAS semantic labeler. This online tagger assigns tags semantics to each lexical item submitted.

4) Lexical valuation/sentiment analysis lists by Hamilton et al. (2016a). Hamilton et al. (2016b) made these lists available on the web. They contain the evaluation of the polarity (positive or negative) of thousands of lexical items, distributed in the decades that they occurred in the corpora surveyed by these authors.

5) Script developed by the supervising, Professor Tony Berber Sardinha. This script, written in Unix Shell, performed all the work of preparing and processing the path to evolve the study. On the procedure part, the survey data comes from the Google Books N-Gram Database. This feature is available free of charge at: http://storage.googleapis.com/books/ngrams/books/datasetsv2.html. An N-gram deals with a sequence of words, such that a unigram is a word, a bigram is a sequence of two words, a trigram, a sequence of three words, a quadrigram, of four words and a pentagram, five words. Google Books offers these five options. On the Google Books N-Gram web address, there are several options of languages for the N-Grams: English, Spanish, Russian, etc. (not currently option for Portuguese). For the English language, there are the following versions:

1) English version 20120701

2) English version 20090703

3) English One Million version 20090705

4) American English version 20120701

5) American English version 20090715

6) British English version 20120701

7) British English version 20090715

8) English Fiction version 20120701

9) English Fiction version 20090715

The Google Books search interface provided by Brigham Young University (through Professor Mark Davies's online corpora project) makes faster and more efficient searches than directly through Google Books Ngram. However, this interface does not provide access to the complete collection "English 20120701", but only American English (items 4 and 5 above) and British English (items 6 and 7 above). As the option for the American database is almost five times greater than that referring to the British database, we opted for the American database, which corresponds to the "American English (155 billion)" option on the interface BYU from Google Books. This option refers to item 5 above, American English version 20090715 and not to item 4, American English version 2012071. Although the help of the BYU interface does not make it clear which of the two bases it used, The results indicate that this is the base of 2009 and not that of 2012 because the results show only until the 2000s; if the database had been used 2012, the results would show the decade of 2010. The 2009 and 2012 versions are different not only in relation to the fact that the 2012 version has three more years of publication, but mainly due to the fact that, of 2012 has considerably increased the number of publications and consequently of words from previous years, as shown in the table below. For example, in the 2009 version, there were just over 3 billion words indexed relating to the year 2005; in the 2012 version, that number increased by more than three times, for more than 10 billion words. The same occurred, in different degrees, in relation to the other years. In the "Total" line, totals from 1810 to 2009 appear. As noted, the base size doubled in size between the 2009 version and the version of 2012. Table 1 shown demonstrates the main comparison between words seems on Google Books and N-grams related to it.

According to Table 1, it is important to note that the names of the databases on the Google website Books N-Gram and the BYU interface reflect the place of publication and not exactly the variety of English. Thus, the publications listed in the "American English" option are not necessarily written in American English, nor do Native American English authors as well as British English write them authors do not reflect purely British English nor British authors. What these bases represent, in fact, they are publications that have been indexed as having been published in USA or Britain. This indexing itself is based on the data recorded by Google Books from the libraries in which the publications were automatically scanned and may contain inaccuracies, as there is no information if the data had a later manual check. Furthermore, even if the registration bibliographic database is reliable, a book published in Great Britain may have been written by an American author and vice versa, putting in question the representativeness of the text as being an example of one of these variants. Thus, we will

Table 1. Comparison between the number of words in the Google Books version 2009 and 2012 N-Grams, American English.

YEAR	2009	2012
2005	3.043.824.240	10.419.437.975
2006	3.124.744.950	10.904.452.060
2007	3.242.955.303	11.401.015.419
2008	2.455.892.145	15.794.843.318
2009	321.421.830	16.545.375.555
TOTAL	**157.388.918.002**	**355.619.887.849**

Source: Enabled by Author.

not speak in terms of "American English" as being the variant studied in this research, but only as "English language". In addition, on the other hand, each of the terms was analyzed using the steps below, which are described below:

1) Search for the term in Brigham's Google Books N-Gram Database interface

Young University immediately preceded by an adjective. We designate the results of this search, in the thesis, with the acronym (adj +). In the development of search, searches were also made with nouns and verbs, but these forms were not incorporated in the final version of the thesis because they scope of the research.

2) Search for the term in Google Books N-Gram Viewer, to be able to view.

Its occurrence and distribution over the studied period (1800-2000).

When relevant, the resulting graph was saved and incorporated into the thesis.

3) Calculation of the normalized frequency of those placed (the adjective associated with search term).

4) Semantic labeling of those placed, using the USAS University of Lancaster. This labeling was used for a first classification of the placed, in order to help the visualization of the semantics of the placed terms. This instrument served as support for the analysis qualitative representation of data. It is necessary to emphasize that semantic tagging does not automatically indicate representations; The identification of representations was made in a qualitative way, using some label categories, but not limited to them.

5) Analysis of the temporal variation of the placed. In this stage, those whose frequency increased and decreased most among the first 50 years (i.e. 1810-1850) and the last 50 years (i.e. 1960-2000) understood by the data. Also identified were those who did not occur in the first 50 years and that had existed in the last 50 years (that is, those that did not necessarily appear in the last 50 years, but which started in 1860).

6) Analysis of the valuation of the placed. In this last stage, those placed were scored on a rating scale, that is, by means of a number that represents its positive or negative charge. Hamilton et al. (2016a) tool has been employed to reach such aim.

4. Research Findings

Results of the Research

This section reports on the results of this research in order to answer the research questions and the final answer to the questions are described in the conclusions section. Further tables, graphs and results can be verified in the full version of the study. Thus, follows the questions to be answered and early settings below:

1) What representations can be identified in relation to the terms surveyed?

2) Is there a difference between the representations of the masculine and feminine terms? Moreover, among children, teenagers and adults?

3) There is a difference between the terms in relation to valuation (positive charge and negative)?

The terms searched were adolescent (s), adult (s), boy (s), child/children, elderly, girl (s), kid (s), man/men, teen (s), teenager (s), woman/women. Below you will be able to find the setting section containing all the frequencies regarding terms use.

Description of Results

In the setting section we want to explain the most frequent semantic category found as well as some other key ones on the study – and here it is found each one of them: A, relative to GENERAL & ABSTRACT TERMS, with 18 placed as "average", "defective", "dependent", "disadvantaged", "exceptional", "good", "hyperactive", "mere", "minor", "natural", "normal", "other", "particular", "perfect", "real", "specific", "true", "typical". The second category is T, on TIME, with 10 placed as "eldest", "modern", "new", "newborn", "old", "older", "oldest", "young", "younger", "youngest". The third category is N, related to NUMBERS & MEASUREMENT, with 10 placed as "additional", "eighth", "fourth", "ninth", "seventh", "single", "tenth", "tiny", "whole". The fourth category is E, relative to EMOTIONAL ACTIONS, STATES & PROCESSES, with 10 placed as "aggressive", "battered", "beloved", "dear", "dearest", "favorite", "happy", "precious", "shy", "unhappy". From these categories, we can suggest the representations of child as the main ones being: 1) evaluative-normativity (beloved, dear, dearest, defective, dependent, exceptional, favorite, good, natural, new, perfect, precious, real, specific, whole); 2) quantification (additional, eighth, eldest, fourth, ninth, seventh, single, tenth); 3) age gradation (minor, newborn, old, older, oldest, young, younger, youngest); 4) behavior (aggressive, happy, shy, unhappy); 5) normativity (average, normal, typical). In short, **"child"** is a figure that is fundamentally evaluated, counted and measured. Regarding the temporal variation, the numbers indicate that both the increase and the decrease are related to a representation of an evaluative nature. However, the increase is related to two aspects of the gradation of the concept of child: small child and older/youngest child, which practically did not exist at the beginning of XIX century. Find below Table 2 prepared aimed to illustrate results:

Table 2. List of adjectives placed immediately to the left of child (adj +) whose normalized frequency per million more grew between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
YOUNG	0.629494	2.316230	1.686736
SMALL	0.051210	1.086450	1.035240
OLD	0.012664	1.040420	1.027756
OLDER	0.011786	0.855754	0.843968
HANDICAPPED	0.001518	0.599538	0.598020
UNBORN	0.053626	0.611776	0.558150
NORMAL	0.000514	0.444128	0.443614
OLDEST	0.048824	0.419234	0.370410
YOUNGEST	0.387324	0.726184	0.338860

Source: Enabled by the Author.

As shown in Table 2, the increase is truly related to two aspects of the gradation of the concept of child: small child and older or youngest child, which practically did not exist at the beginning of XIX century. Table 3 details for analysis the list of adjective in the right side of the term child (adj +).

As shown in Table 3, it is possible to understand the list of adjectives placed immediately to the right of child (adj +) whose normalized frequency per million more decreased between 1810-1850 and 1960-2000. Table 4 shows the positions that emerged between these two periods. These places indicate the emergence of representations related to health (retarded child, autistic child, disabled child, exceptional child). Therefore, there was a rise in the child's medical representation.

According to Table 4 described, it shows the positions that emerged between these two periods. These places indicate the emergence of representations related to health. It suggests that the concept of "child" has changed from being monolithic

Table 3. List of adjectives placed immediately to the right of child (adj +) whose normalized frequency per million more decreased between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
POOR	1.253860	0.362834	−0.891026
BELOVED	0.894838	0.154844	−0.739994
MERE	0.284836	0.072782	−0.212054
SWEET	0.249676	0.055336	−0.194340
LOVELY	0.202118	0.042296	−0.159822
FAIR	0.162728	0.012256	−0.150472
INNOCENT	0.222098	0.126074	−0.096024
DEAREST	0.109442	0.018924	−0.090518
BEAUTIFUL	0.192352	0.111660	−0.080692
GOOD	0.222136	0.148272	−0.073864

Source: Enabled by the Author.

Table 4. List of adjectives placed immediately to the left of child (adj +) that appeared between 1810-1850 and 1960-2000 (per million words).

Collocate	1810-1850	1960-2000	Difference
RETARDED	0.000000	0.592912	0.592912
PRESCHOOL	0.000000	0.313574	0.313574
DISABLED	0.000000	0.250116	0.250116
AUTISTIC	0.000000	0.230778	0.230778
BATTERED	0.000000	0.170482	0.170482
HYPERACTIVE	0.000000	0.142698	0.142698
INNER	0.000000	0.127150	0.127150
DISADVANTAGED	0.000000	0.108272	0.108272
EXCEPTIONAL	0.000000	0.101380	0.101380
DEFECTIVE	0.000000	0.101102	0.101102

Source: Enabled by the Author.

to somewhat gradual in terms of size and/or age. There was also an increase of the concept of a child related to abortion (unborn child) and physical issues (handicapped child). On the other hand, there was a decrease in the representation of child as an innocent child, like "innocent child", "beloved child", "sweet child". Even "poor child" refers to a value judgment not always related to financial condition. The results reflect those of "Child", insofar as the placements whose frequency has increased comprise gradation (Young child, younger child, older child) and health aspects (handicapped, retarded). However, a racial dimension has also emerged in the representation of "children", which was not apparent as "child" (black children, white children).

On the other hand, **"children"** is a term more willing to be associated with racial aspects than "child" is. At the same time, those who became most rare in the comparison are those focused on the idea of innocence and purity (beloved children, good children, lovely children, innocent children), including religious aspects (spiritual children). Interesting to observe "fatherless children", which had its frequency greatly decreased in the recent period, this suggests that "children" are no longer characterized by the absence of the father. In relation to the placements that appeared in the comparison of the two periods, there is clearly a medical-clinical representation, through "autistic children", "disabled children", "psychotic children", etc. There is also the emergence of a representation of social problems (disadvantaged children).

In short, "children" is a concept represented essentially in an evaluative way, descriptive and clinical. The view of **"children"** as innocent and pure beings, who existed at the beginning of the last century, gave rise to a more gradual, racial, medical and social. The most frequent semantic category is O, relative to SUBSTANCES, MATERIALS, OBJECTS AND EQUIPMENT, with 16 placed as "attractive", "beautiful", "black", "blond", "blonde", "bright", "brown", "charming", "handsome", "lovely", "nice", "pale", "prettiest", "smart", "white", "wretched". The second category is A, relative GENERAL & ABSTRACT TERMS, with 16 placed as "average", "bad", "best", "decent", "fine", "good", "honest", "lucky", "mere", "other", "private", "simple", "slender", "strange", "unfortunate", "wonderful". The third category is T, relating to TIME, with 11 placed as "adolescent", "eldest", "modern", "new", "old", "older", "oldest", "teenage", "young", "younger", "youngest". The fourth category is S, relative to SOCIAL ACTIONS, STATES & PROCESSES, with 7 placed as "catholic", "Christian", "foolish", "Jewish", "sensible", "silly" and "unmarried". Table 5 and Table 6 are demonstrating the main results found about this term:

Table 5 described the list of adjectives placed immediately to the left of children (adj +) whose normalized frequency per million more grew between 1810-1850 and 1960-2000. Therefore, Table 6 to be shown demonstrates the list of adjectives placed immediately to the right of children (adj +) whose normalized frequency per million more decreased between 1810-1850 and 1960-2000.

As it is shown, Table 6, details the list of adjectives placed immediately to the right of children (adj +) whose normalized frequency per million more

Table 5. List of adjectives placed immediately to the left of children (adj +) whose normalized frequency per million more grew between 1810-1850 and 1960-2000.

Collocate	1800-1850	1960-2000	Difference
YOUNG	1.564050	8.390960	6.826910
OTHER	1.402700	5.499820	4.097120
OLDER	0.106520	3.238930	3.132410
SMALL	0.381808	2.181450	1.799642
HANDICAPPED	0.001214	1.421370	1.420156
BLACK	0.041190	1.421060	1.379870
RETARDED	0.000744	1.300160	1.299416
YOUNGER	0.583920	1.876730	1.292810
AMERICAN	0.038722	1.256570	1.217848
WHITE	0.113710	0.935450	0.821740

Source: Enabled by the Author.

Table 6. List of adjectives placed immediately to the right of children (adj +) whose normalized frequency per million more decreased between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
BELOVED	0.476306	0.122270	−0.354036
POOR	0.956124	0.647954	−0.308170
FATHERLESS	0.245020	0.050820	−0.194200
SPIRITUAL	0.219106	0.034776	−0.184330
HELPLESS	0.193356	0.050210	−0.143146
GOOD	0.213982	0.098348	−0.115634
LOVELY	0.128326	0.042970	−0.085356
NATURAL	0.153356	0.072068	−0.081288
MERE	0.100614	0.031454	−0.069160
INNOCENT	0.184718	0.138276	−0.046442

Source: Enabled by the Author.

decreased between 1810-1850 and 1960-2000. The fifth category is E, relative to EMOTIONAL ACTIONS, STATES & PROCESSES, with 7 placed as "brave", "dearest", "gentle", "happy", "hearted", "popular" and "unhappy". The sixth category is N, relating to NUMBERS & MEASUREMENT, with 6 placed as "big", "single", "small", "tall", "thin" and "tiny". The seventh category is B, relating to THE BODY & THE INDIVIDUAL, with 5 placed as "haired", "healthy", "naked", "pregnant" and "sick".

The following representations seem to be related to **"girl"**: 1) evaluative (bad, brave, charming, dearest, decent, fine, foolish, gentle, good, hearted, honest, lovely, lucky, new, sensible, silly, simple, smart, unfortunate, wonderful, wretched), 2) physics (attractive, beautiful, big, blond, blonde, haired, handsome, pale, prettiest, slender, small, tall, thin, tiny), 3) age (adolescent, old, older, oldest, teenage, young, younger, youngest), 4) spirituality (catholic, Christian, Jewish), 5) civil (single, unmarried). Attention is drawn to the joint representation of physical appearance and evaluation-normativity, for being quite numerous, such as bigrams like "attractive girl", "beautiful girl", "blond/e girl", "thin girl", "tall girl", etc. Therefore, we can suggest that the fundamental representation of "girl" is of a physical-evaluative nature.

The most frequent semantic category is A, relative to GENERAL & ABSTRACT TERMS, with 16 placed as "average", "bad", "best", "fine", "good", "great", "idle", "mere", "normal", "ordinary", "other", "private", "real", "stable", "strange", "wonderful". The second category is O, relative to SUBSTANCES, MATERIALS, OBJECTS AND EQUIPMENT, with 13 placed as "beardless", "beautiful", "black", "blond", "bright", "golden", "handsome", "lovely", "nice", "ragged", "smart", "white", "wretched". The third category is T, relative to TIME, with 10 placed as "adolescent", "eldest", "new", "old", "older", "oldest", "teenage", "young", "younger", "youngest". The fourth category is X, relating to PSYCHOLOGICAL ACTIONS, STATES & PROCESSES, with 8 placed as "active", "blind", "clever", "dull", "intelligent", "quiet", "sensitive", "sweet". The fifth category is N, relative to NUMBERS & MEASUREMENT, with 7 placed as "big", "large", "small", "smaller", "smallest", "tall", "tiny".

These categories indicate the following possible representations: 1) evaluative normativity, 2) physical appearance, 3) age range and 4) ability intellectual. Below it can be found the list of adjectives for the term. Table 7 is describing the immediate terms found on the left of the term *girl*:

According to data shown, the Table 7 has demonstrated the list of adjectives placed immediately to the left of girl (adj +) whose frequency normalized per

Table 7. List of adjectives placed immediately to the left of girl (adj +) whose frequency normalized per million more grew between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
OLD	0.227428	2.603330	2.375902
YOUNG	0.171872	0.353500	1.181628
SMALL	0.153728	0.972816	0.819088
WHITE	0.033370	0.262906	0.229536
BIG	0.038124	0.250232	0.212108
OLDER	0.009794	0.189628	0.179834
AMERICAN	0.014992	0.187816	0.172824
JEWISH	0.005004	0.151964	0.146960
ADOLESCENT	0.000134	0.146366	0.146232
BAD	0.094516	0.234444	0.139928

Source: Enabled by the Author.

million more grew between 1810-1850 and 1960-2000 and Table 8 demonstrates the list of adjectives placed immediately to the right of girl (adj +) whose frequency normalized per million more decreased between 1810-1850 and 1960-2008.

As is described, Table 8 details the list of adjectives placed immediately to the right of **girl (adj +)** whose frequency normalized per million more decreased between 1810-1850 and 1960-2008 and Table 9 shows the list of adjectives placed immediately to the left of boy (adj +) whose normalized frequency per million more grew between 1810-1850 and 1960-2000.

As demonstrated, Table 9 shows the list of adjectives placed immediately to the left of boy (adj +) whose normalized frequency per million more grew between 1810-1850 and 1960-2000. As with the other terms discussed above, there is a great deal of evaluative component linked to the concept of **"boy"**, which undergoes judgment based on in physical appearance and intellectual ability.

Table 8. List of adjectives placed immediately to the right of **girl (adj +)** whose frequency normalized per million more decreased between 1810-1850 and 1960-2008.

Collocate	1810-1850	1960-2000	Difference
OLD	0.055030	1.948580	1.893550
YOUNG	1.977080	3.315300	1.338220
WHITE	0.022816	0.315108	0.292292
AMERICAN	0.034842	0.311230	0.276388
OTHER	0.045034	0.305208	0.260174
NICE	0.025778	0.285194	0.259416
TEENAGE	0.000606	0.219758	0.219152

Source: Enabled by the Author.

Table 9. List of adjectives placed immediately to the left of boy (adj +) whose normalized frequency per million more grew between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
OLD	0.227428	2.603330	2.375902
YOUNG	0.171872	1.353500	1.181628
SMALL	0.153728	0.972816	0.819088
WHITE	0.033370	0.262906	0.229536
BIG	0.038124	0.250232	0.212108
OLDER	0.009794	0.189628	0.179834
AMERICAN	0.014992	0.187816	0.172824
JEWISH	0.005004	0.151964	0.146960
ADOLESCENT	0.000134	0.146366	0.146232
BAD	0.094516	0.234444	0.139928

Source: Enabled by the Author.

There is also a component of gradation of this concept, in which the idea of "boy" is dissected in nuances old and size. In relation to the temporal variation, those whose frequency grew the most reflect representations geared to age (old boy, older boy), appearance physical (small boy), which in turn is also related to age, ethnicity (white boy), judgment of value (bad boy), nationality (American boy) and spirituality (Jewish boy). In turn, the representations that have declined the most over time are related to ideals like "brave boy" and "noble boy" as well as value judgments (idle boy) and judgment (lovely boy). Table 10 describe the main results about the adjectives on the term boy:

As described, Table 10 shows a list of adjectives placed immediately to the left of boy (adj +) whose frequency normalized per million more decreased between 1810-1850 and 1960-2000. The terms analyzed in relation to **adolescence** are not marked by **gender**, because as we saw in the analysis presented above, the gender marking occurs with the junction of the term referring to adolescence to the generic term (e.g. teenage girls). Regarding the ***temporal variation*** of the term itself (without participating in the bigrams), the graph (Figure 1) shows that "teen" and "teens" appeared first; however, searches in the texts revealed that it is not a reference to adolescence, but rather a "ten" (e.g. thirteen). With the sense of "adolescent", the terms emerged in the early twentieth century. The most frequent term is adolescent (s). The graph shows a gradual increase in the use of these terms, differently (with the exception of "teen") in the 20th century. Figure 1 shows the variation mentioned.

Regarding the term teenager for example, The most frequent category is A, relative to GENERAL & ABSTRACT TERMS, with 14 placed as "average", "awkward", "dependent", "difficult", "mere", "normal", "ordinary", "other" , "private", "real", "regular", "runaway", "typical", "unruly". The second category is E, relative to EMOTIONAL ACTIONS, STATES & PROCESSES, with 12

Table 10. List of adjectives placed immediately to the left of boy (adj +) whose frequency normalized per million more decreased between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
POOR	0.940344	0.405482	−0.534862
ELDEST	0.228362	0.079018	−0.149344
MERE	0.203820	0.071412	−0.132408
BRAVE	0.148014	0.033030	−0.114984
LOVELY	0.139582	0.028638	−0.110944
NOBLE	0.117594	0.009984	−0.107610
HEARTED	0.120332	0.017738	−0.102594
FINE	0.163972	0.074524	−0.089448
HAPPY	0.101970	0.030150	−0.071820
IDLE	0.076098	0.005680	−0.070418

Source: Enabled by the Author.

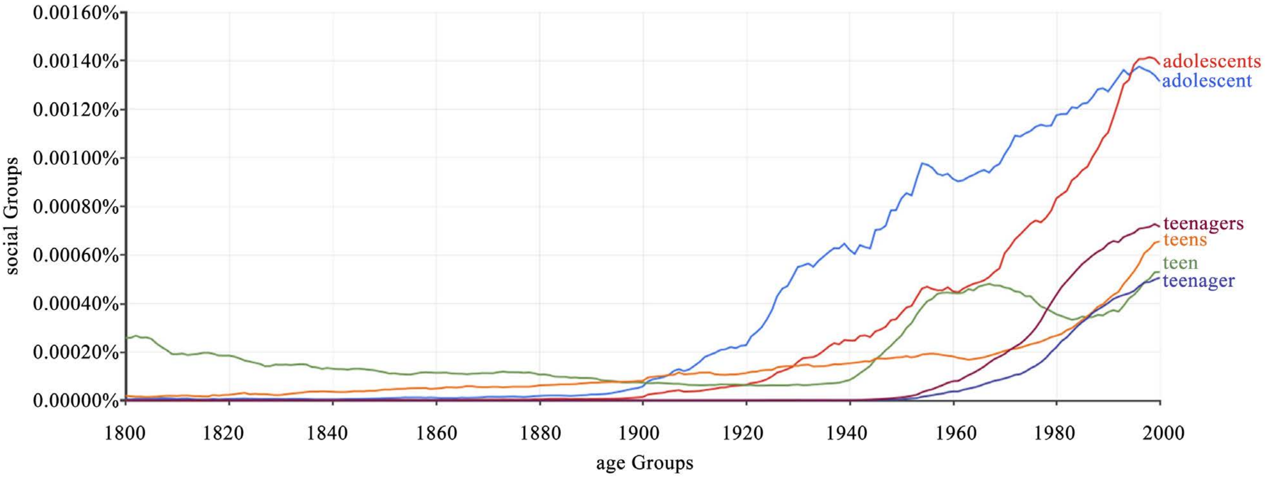

Figure 1. Time variation of terms related to adolescence age group. Source: Enabled by Google Books.

placed as "angry", "depressed", "happy", "insecure", "moody", "nervous", "popular", "scared", "shy", "sullen", "troubled", "unhappy". The third category is T, relative to TIME, with 8 placed as "early", "late", "mature", "modern", "new", "old", "young", "younger". The fourth category is X, relative to PSYCHOLOGICAL ACTIONS, STATES & PROCESSES, with 6 placed as "active", "bored", "conscious", "deaf", "sensitive", "talented".

It has been possible to detect the following possible representations: 1) evaluative-normativity/normativity, 2) mental aspects and 3) age gradation. The evaluative/normative representation includes "average teenager", "difficult teenager", "ordinary teenager", etc. The representation of mental states incorporates emotional aspects such as "angry teenager", "nervous teenager", "bored teenager" and "unhappy teenager" but also clinicians like "depressed teenager". It is interesting how many of these states are problematic, such as "scared teenager", "troubled teenager" and "insecure teenager". The representation of age grading includes "early teenager", "late teenager" and "young teenager", again showing that the concept of adolescence is elastic. Thus, in general, the representations of "teenager" build an unfavorable scenario for adolescence, such as the locus of behavioral, psychological and personality issues. Regarding the temporal variation, the bigrams that emerged more recently reflect evaluative representations based on social issues, such as "pregnant teenager", "rebellious teenager" and "unmarried teenager", in addition to ethnic (black teenager) and national (American) issues teenager). There was no occurrence of bigrams whose frequency decreased. Table 11 shows the list of adjectives placed immediately to the left of teenager (adj +) that appeared between 1810-1850 and 1960-2000 (per million words).

According to mentioned, Table 11 details as said, the list of adjectives placed immediately to the left of teenager (adj +) that appeared between 1810-1850 and 1960-2000 (per million words). Regarding the term **adult**, the most frequent semantic category is A, relative to GENERAL & ABSTRACT TERMS, with 30 placed as "appropriate", "average", "certain", "civilized", "common", "conventional",

Table 11. List of adjectives placed immediately to the left of teenager (adj +) that appeared between 1810-1850 and 1960-2000 (per million words).

Collocate	1810-1850	1960-2000	Difference
YOUNG	0.000000	0.055610	0.055610
PREGNANT	0.000000	0.027250	0.027250
AMERICAN	0.000000	0.026324	0.026324
BLACK	0.000000	0.024424	0.024424
TYPICAL	0.000000	0.011922	0.011922
REBELLIOUS	0.000000	0.011158	0.011158
AVERAGE	0.000000	0.009834	0.009834
NORMAL	0.000000	0.008990	0.008990
MALE	0.000000	0.008968	0.008968
UNMARRIED	0.000000	0.007290	0.007290

Source: Enabled by the Author.

"dependent", "different", "familiar", "important", "liberal", "major", "normal", "ordinary", "other", "particular", "positive", "real", "regular", "responsible", "serious", "significant", "special", "specific", "stable", "strange", "superior", "typical", "usual", "various". The second category is T, relative to TIME, with 16 placed as "chronic", "complete", "contemporary", "current", "daily", "early", "elderly", "late", "later", "mature", "modern", "new", "old", "older", "young", "younger". The third category is N, relative to NUMBERS & MEASUREMENT, with 13 placed as "additional", "adequate", "entire", "full", "high", "higher", "large", "miniature", "obese", "single", "small", "subsequent", "whole". The representations associated with "adult" appear to be as follows: 1) evaluative-normative character, e.g. "average adult", "familiar adult", "typical adult"; and 2) age range, e.g. "mature adult", "young adult", "old adult", "elderly adult". The representation of adult life through normativity is quite expressive, being reflected in many bigrams. The gradation of the adult phase is less expressive, unlike adolescence, shown in the sections above, which is quite graded in nuances of time. In fact, the range of representations of adult life through this term is restricted, compared to other terms analyzed here.

Other minority representations are of a political-ideological character, e.g. "liberal adult" and physical appearance, "obese adult". Regarding the temporal variation, the bigrams that became more frequent in the most recent period (Table 12) point to representations already identified, such as the age gradation (young adult, early adult, mature adult), to the evaluative-normativity (average adult, responsible adult), health (healthy adult) and gender (male adult). Among these, the biggest highlight is "young adult", which had an expressive growth compared to the other bigrams. This indicates the widening of the spectrum of adult life, incorporating a "young" phase, which was not common in the 19th century. There was no decrease in frequency. Table 12 demonstrates the main

Table 12. List of adjectives placed immediately to the left of adult (adj +) whose frequency normalized per million more grew between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
YOUNG	0.006236	1.570280	1.564044
OTHER	0.004758	0.261472	0.256714
AVERAGE	0.000408	0.242550	0.242142
EARLY	0.001086	0.200656	0.199570
HEALTHY	0.024108	0.186714	0.162606

Source: Enabled by the Author.

findings.

As mentioned, Table 12 brings up the List of adjectives placed immediately to the left of adult (adj +) whose frequency normalized per million more grew between 1810-1850 and 1960-2000.

The representation of the ***adult phase age group*** as a larger spectrum of life is confirmed in the list of terms that emerged in the last half-century, in which "older adult" is the most expressive; "Younger adult" also fits this representation. Other representations include that of competence (competent adult), generality (universal adult) and clinical (retarded adult). Regarding the term woman, the most frequent semantic category is A, relative to GENERAL & ABSTRACT TERMS, with 21 placed as "average", "bad", " certain ", "decent", "excellent", "extraordinary", "fine", "good", "great", "honest", "ideal", "ordinary", "other", "particular", "perfect", "real", "remarkable", "strange", "true", "unfortunate", "wonderful".

The second category is S, relative to SOCIAL ACTIONS, STATES & PROCESSES, with 12 placed as "Christian", "foolish", "free", "independent", "Jewish", "pious", "religious", "sensible", "strong", "unmarried", "weak", "wise". The third category is O, relating to SUBSTANCES, MATERIALS, OBJECTS AND EQUIPMENT, with 11 placed as "attractive", "beautiful", "black", "blond", "charming", "handsome", "lovely", "nice", "stout", "white", "wretched". The fourth category is T, relative to TIME, with 8 placed as "elderly", "mature", "modern", "new", "old", "older", "young", "younger".

The main representations reflected in these categories seem to include: 1) evaluative-normativity, e.g. "average woman", "fine woman", "remarkable woman"; 2) spirituality, e.g. "Christian woman", "Jewish woman"; 3) physical appearance, e.g. "attractive woman", "charming woman", "blond woman"; and 4) age group, e.g. "elderly woman", "mature woman", "old woman". Thus, the representation of "woman" seems to be, in general, of a physical-evaluative-spiritual nature. Regarding the temporal variation, the representations that grew the most between the periods compared were the age (young woman, old woman, elderly woman), the ethnic (black woman, white woman), the gestational age (pregnant women), aesthetics (beautiful woman), and nationality (American woman). In turn, the ones that most fell back in terms of frequency were the evaluative and/or social class representations (poor woman, excellent woman, good woman, and lovely woman), idealized (virtuous woman) and disposition/humor (happy woman, unhappy woman). In other words, there was a decrease in subjective, idealized and humor representations and an increase in age, ethnic and aesthetic representations. Table 13 demonstrates exactly such a perspective in numbers:

As I said, Table 13 listed adjectives placed immediately to the left of woman (adj +) whose normalized frequency per million more grew between 1810-1850 and 1960-2000. Table 14 illustrates the adjectives placed to the left.

According to results above, Table 14 lists adjectives placed immediately to the left of woman (adj +) whose frequency normalized per million more decreased between 1810-1850 and 1960-2000.The analysis identified that there were no bigrams that emerged between 1810-1850 and 1960-2000. Representations

Table 13. List of adjectives placed immediately to the left of woman (adj +) whose normalized frequency per million more grew between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
YOUNG	4.301300	7.443890	3.142590
OLD	5.497140	7.015630	1.518490
BLACK	0.117678	1.377330	1.259652
PREGNANT	0.083362	1.337490	1.254128
WHITE	0.169994	1.119510	0.949516
OLDER	0.005764	0.916570	0.910806
AMERICAN	0.068318	0.855820	0.787502
OTHER	0.289500	0.986874	0.697374
BEAUTIFUL	0.680454	1.091720	0.411266
ELDERLY	0.122044	0.483076	0.361032

Source: Enabled by the Author.

Table 14. List of adjectives placed immediately to the left of woman (adj +) whose frequency normalized per million more decreased between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
POOR	1.987660	0.572376	−1.415284
GOOD	1.353780	0.534862	−0.818918
EXCELLENT	0.307740	0.026572	−0.281168
VIRTUOUS	0.360470	0.090978	−0.269492
HAPPY	0.312920	0.103326	−0.209594
HEARTED	0.274746	0.078532	−0.196214
UNHAPPY	0.238160	0.057764	−0.180396
PIOUS	0.202336	0.037056	−0.165280
LOVELY	0.301020	0.143134	−0.157886
FINE	0.249938	0.098802	−0.151136

Source: Enabled by the Author.

associated with "women" mainly include the following: 1) evaluative-normativity: "good women", "great women", "normal women"; 2) spirituality: "Catholic women", "Christian women", "holy women", "Jewish women", "Muslim women"; 3) age group: "mature women", "old women", "young women"; 4) origin: "foreign women", "immigrant women", "native women"; 5) medical or clinical aspects: "diabetic women", "disabled women", "infertile women"; 6) body or eroticism: "naked women"; 7) temporality: "contemporary women", "modern women"; and 8) gender identification: "bisexual women". In short, the representation of "women" seems to be based on age, body and spirituality. Further details regarding the missing settings can be found in the full study and added in the current manuscript upon demand. With this disconnection, it became necessary to specify, when precise, the marital status. Among the representations that have become rarer are old age (old women), idealization (virtuous women), evaluation (helpless women), spirituality (holy women), appearance (fair women), and origin (Turkish) women, roman women). Below is the illustration of the main findings seen in Table 15.

Thus, Table 15 has listed adjectives placed immediately to the left of women (adj +) whose frequency normalized per million more grew between 1810-1850 and 1960-2000. Table 16 is describing the adjectives placed to the left:

As mentioned, Table 16 listed adjectives placed immediately to the left of women (adj +) whose normalized frequency per million more decreased between 1810-1850 and 1960-2000. In respect to the bigrams that emerged in the last surveyed period, the bigrams point to representations of origin (immigrant women, Hispanic women, and urban women), intimacy (menopausal women), gender identification (heterosexual women, lesbian women) and clinic (diabetic women, obese women). In respect to the term **man**, the most frequent semantic category is A, relative to GENERAL & ABSTRACT TERMS, with 29 placed as

Table 15. List of adjectives placed immediately to the left of women (adj +) whose frequency normalized per million more grew between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
OTHER	0.710388	4.506230	3.795842
BLACK	0.039312	3.487860	3.448548
PREGNANT	0.193678	3.097960	2.904282
YOUNG	2.089380	4.820880	2.731500
AMERICAN	0.171262	2.775110	2.603848
WHITE	0.095134	2.135660	2.040526
OLDER	0.003844	1.849450	1.845606
MARRIED	0.260744	1.056550	0.795806
SINGLE	0.096436	0.845206	0.748770
UNMARRIED	0.224382	0.905354	0.680972

Source: Enabled by the Author.

Table 16. List of adjectives placed immediately to the left of women (adj +) whose normalized frequency per million more decreased between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
OLD	1.215040	1.146330	−0.068710
HOLY	0.127488	0.061598	−0.065890
HELPLESS	0.099480	0.034080	−0.065400
TURKISH	0.091092	0.039008	−0.052084
FAIR	0.075416	0.029434	−0.045982
GOOD	0.174596	0.143122	−0.031474
ROMAN	0.086880	0.057108	−0.029772
VIRTUOUS	0.066534	0.040408	−0.026126
LOVELY	0.061080	0.042426	−0.018654

Source: Enabled by the Author.

"average", "bad", "best", "busy", "certain", "civilized", "common", "different", "excellent", "extraordinary", "fine", "good", "great", "greatest", "honest", "mere", "natural", "normal", "ordinary", "other", "perfect", "practical", "prudent", "real", "remarkable", "strange", "true", "unfortunate", "wonderful". The second category is S, relative to SOCIAL ACTIONS, STATES & PROCESSES, with 15 placed as "Christian", "fellow", "free", "gay", "holy", "pious", "powerful", "public", "reasonable", "religious", "sensible", "spiritual", "strong", "wise", "worthy". The third category is T, relative to TIME, with 8 placed as "elderly", "modern", "new", "old", "older", "primitive", "young", "younger". Representations emanating from these categories appear to include the following: 1) Evaluative-normativity, e.g. "average man", "common man", "normal man"; 2) character, e.g. "honest man", "prudent man", "true man", "honest man"; 3) Spirituality, e.g. "Christian man", "holy man", "spiritual man"; 4) Strength, e.g. "strong man"; 5) Wisdom, e.g. "wise man", "reasonable man", "sensible man"; and 6) Gender identification, e.g. "gay man". Regarding the temporal variation, bigrams point to the following representations that would have grown in relation to the last surveyed period: 1) ethnicity, e.g. "black man", "white man"; 2) time gradation, e.g. "older man", "younger man"; 3) temporality, e.g. "modern man"; 4) evaluative-normativity, e.g. "common man", "average man", "primitive man"; 5) appearance, e.g. "haired man". In turn, bigrams whose frequency has decreased include the following representations: 1) evaluative-normativity, e.g. "good man", "great man"; 2) age gradation, e.g. "young man", "old man"; 3) social class, e.g. "rich man", "poor man"; 4) character/wisdom, e.g. "honest man", "wise man"; and 5) naturalness, e.g. "natural man". Table 17, lists adjectives placed immediately to the left of man (adj +) whose normalized frequency per million more grew between 1810-1850 and 1960-2000.

Therefore, Table 17 lists adjectives placed immediately to the left of man (adj

+) whose normalized frequency per million more grew between 1810-1850 and 1960-2000. Table 18 shows the adjective placed to the left of the term man.

As mentioned, Table 18 shows a list of adjectives placed immediately to the left of man (adj +) whose normalized frequency per million more decreased between 1810-1850 and 1960-2000. The plural form of the term analysis can be found in the thesis of the current research. In order to accomplish with the number of words allowed in this manuscript, further data can be found in extra attachment added in the submission. In this next analysis, the positions of the female terms were compared with those of the male terms. The feminine terms are girl, girls, woman and women, the masculine ones: boy, boys, man and men. The comparison was made using a script developed by the tutor. There were 120

Table 17. List of adjectives placed immediately to the left of man (adj +) whose normalized frequency per million more grew between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
BLACK	0.627870	2.783200	2.155330
WHITE	3.362120	5.077600	1.715480
MODERN	0.024958	1.254180	1.229222
OLDER	0.075632	1.263070	1.187438
BIG	0.057732	1.072200	1.014468
COMMON	0.367278	1.035050	0.667772
AVERAGE	0.006292	0.658984	0.652692
YOUNGER	0.176650	0.720292	0.543642
PRIMITIVE	0.034670	0.475964	0.441294
HAIRED	0.103794	0.477008	0.373214

Source: Enabled by the Author.

Table 18. List of adjectives placed immediately to the left of man (adj +) whose normalized frequency per million more decreased between 1810-1850 and 1960-2000.

Collocate	1810-1850	1960-2000	Difference
GOOD	9.89830	2.582820	−7.237010
YOUNG	25.768000	20.332200	−5.435800
GREAT	6.367880	1.754600	−4.613280
POOR	5.329470	1.662280	−3.667190
HONEST	4.401220	0.815210	−3.586010
WISE	4.524510	1.267420	−3.257090
RICH	3.544000	1.337060	−2.206940
OLD	20.445300	18.363500	−2.081800
NATURAL	1.992950	0.387134	−1.605816
HAPPY	1.773030	0.395772	−1.377258

Source: Enabled by the Author.

singular places (i.e., not repeated) in the bigrams compared according to gender. Of these, 31 occurred only in the female bigrams, 51 only in the male ones and 38 occurred in both groups. This indicates that although there is a specialization of those placed, one-third of them do not distinguish between one gender and the other (N = 38, 31.7%). Table 19 also will bring the collocates resulting from the comparison of bigrams based on gender.

As I said, Table 19 highlights the collocates resulting from the comparison of bigrams based on gender. The results indicate that those placed in the female terms mainly reflect location (foreign, immigrant, local, native, rural, urban), physical condition (attractive, beautiful, fair, obese, diabetic, disabled), reproduction (childless, infertile, pregnant), virtuosity (respectable, virtuous), spirituality (catholic, Muslim), marital status (married, widowed) and gender identification (bisexual, lesbian). In turn, the male bigram placements reflect a large component of representation of superiority/success, with those placed as ambitious, best, brave, civilized, distinguished, eminent, greatest, honest, illustrious, important, influential, mighty, powerful, principal, reasonable, remarkable, sensible, thoughtful, true, wise, wisest and worthy. There is also a representation of the male through occupation (literary, medical, military, public, scientific), financial success (rich, richest, wealthy), physical appearance (big, red), gender identification (gay), kindness (pious), warmongering (armed), among others. As can be seen, there is a very big difference between the representations of the female and male human beings in the data. The feminine is generally represented from the point of view of physical appearance, its origin/location, reproductive (in-) capacity, idealized virtuosity, spirituality.

The masculine, however, appears generally represented as superior, successful,

Table 19. Positions resulting from the comparison of bigrams based on gender.

Bigrams of terms	Collocated	N
Female only	*attractive, battered, beautiful, bisexual, catholic, childless, contemporary, diabetic, disabled, fair, foreign, helpless, homeless, immigrant, independent, infertile, lesbian, local, lovely, married, mature, Muslim, naked, native, obese, pregnant, respectable, rural, urban, virtuous, widowed*	31
Male only	*able, ablest, ambitious, armed, best, big, blind, brave, civilized, common, dead, desperate, distinguished, eminent, fellow, gay, greatest, homosexual, honest, illustrious, important, influential, lesser, literary, medical, mighty, military, mortal, pious, powerful, practical, primitive, principal, public, reasonable, red, remarkable, rich, richest, scientific, sensible, sick, strange, thoughtful, true, wealthy, wicked, wild, wise, wisest, worthy*	51
Both genders	*active, bad, black, certain, Christian, different, elderly, famous, free, good, great, healthy, heterosexual, holy, innocent, intelligent, Jewish, modern, new, noble, normal, old, older, ordinary, other, poor, professional, prominent, real, religious, single, southern, strong, successful, unmarried, white, young, younger*	38

Source: Enabled by the Author.

linked to physical or intellectual work. At this stage of the analysis, the terms related to childhood, those related to adolescence and those related to adult life were compared. The terms related to childhood are boy, boys, child, children, girl, girls, kid and kids; those related to adolescence are adolescent, adolescents, teen, teenager and teenagers. Finally, those related to adult life are adult, adults, man, men, woman and women. The comparison was made using a script developed by the tutor. A total of 150 places of the bigrams were computed compared according to the age group. Of these, 40 occurred only in the bigrams related to childhood, 43 only in those related to adolescence, 53 only in those related to adulthood and 14 occurred in the three groups. This indicates that there is a marked specialization of those placed, as only a small number (9.3%) occurs in the three categories, unlike bigrams related to gender. Table 20 brings the result of comparison of bigram's based on age groups.

As written and fully descrived, Table 20 has brought in the research the result of comparison of bigram's based on age groups. The results of the representation analysis show three dramatically different patterns of the three groups. In childhood, physical representations are predominant (barefoot, beardless, beautiful, bigger, gallant, handsome, large, larger, small, and smaller). There are also representations of behavior (aggressive, happy, idle, merry, nice, mischievous, rough, rude, wanton), own age (oldest, preschool, senior, teenage, youngest) and superiority/virtue (popular, promising, smart, stable). In adolescence, the

Table 20. Placed as a result of comparing bigram's based on age group.

Bigrams of terms	*Collocates*	
Infancy only	*adolescent, aggressive, barefoot, beardless, beautiful, bigger, bright, catholic, clever, fine, gallant, handsome, happy, hungry, hyperactive, idle, large, larger, mere, merry, mischievous, naked, native, naughty, nice, oldest, popular, preschool, promising, ragged, rough, rude, senior, small, smaller, smart, stable, teenage, wanton, youngest*	0
Adolescence only	*affluent, angry, bisexual, bored, contemporary, deaf, depressed, difficult, disabled, disadvantaged, drunk, drunken, early, female, handicapped, ill, immigrant, impressionable, inexperienced, involved, late, lesbian, male, mature, noisy, obese, overweight, pregnant, rebellious, restless, rowdy, runaway, suburban, suicidal, sullen, talented, troubled, typical, unemployed, unwed, urban, violent, vulnerable*	43
Adult life only	*able, ablest, ambitious, armed, blind, civilized, common, desperate, distinguished, elderly, eminent, famous, free, great, greatest, holy, honest, illustrious, important, influential, innocent, lesser, literary, medical, mighty, military, mortal, pious, powerful, practical, primitive, principal, professional, prominent, public, reasonable, red, religious, remarkable, richest, scientific, sensible, sick, strange, strong, successful, thoughtful, true, wealthy, wicked, wise, wisest, worthy*	53
All stages of life	*active, black, Christian, gay, Jewish, normal, old, older, other, poor, real, white, young, younger*	14

Source: Enabled by the Author.

main representations are behavioral (angry, bored, noisy, restless, rowdy, sullen, violent), social (disadvantaged, drunk, drunken, runaway, unemployed, vulnerable), clinics (deaf, depressed, disabled, handicapped, ill, suicidal), gender identification (bisexual, female, lesbian, male) and inferiority (impressionable, inexperienced, troubled). In adulthood, the dominant representation is superiority/success (ambitious, civilized, distinguished, eminent, famous, great, greatest, honest, illustrious, important, influential, mighty, powerful, principal, prominent, reasonable, remarkable, sensible, successful, thoughtful, true, wise, wisest, worthy), followed by representations of occupation (literary, medical, military, professional, public, scientific) and inferiority (desperate, lesser, primitive, strange, wicked). Thus, based on these categories, there seems to be a pattern of representation of the three phases that follows a trajectory that goes from physical appearance, behavior and age classification, in childhood, to issues behavioral, clinical, gender identification and inferiority, in adolescence, for a representation of success and superiority, in adult life. There is, therefore, strong evidence of an appreciation of childhood and especially adult life, to the detriment of adolescence.

5. Conclusion

This research aimed at identifying the representations associated with terms that designate as Human Being in English Language, from the use of Google Books NGrams database online available, covering a period ranging from beginning of the 19th century to the beginning of the 21st century. A total of twenty terms were investigated, divided between terms related to childhood, feminine (girl, girls), male (boy, boys) and unmarked by gender (child, children, kid, kids); terms related to adolescence (all not marked by gender, adolescent, adolescents, teen, teens, teenager, teenagers) and adulthood, female (woman, women), masculine (man, men) and not marked by gender (adult, adults). The research questions asked were: 1) what representations can be identified in relation to the terms surveyed. 2) Is there a difference between the representations of the masculine and feminine terms? Moreover, among children, teenagers and adults? 3) There is a difference between the terms in relation to valuation (positive charge and negative)? In order to answer the research questions, tools were used computationally, availably on the network and developed specifically for this research. In addition, an interpretative-qualitative analysis of the data was carried out, because no tool is able to extract representations automatically from satisfactory way. Regarding the first research question, the results showed a wide range of representations (more than 30). The evaluative representation of being human was the most constant among the terms, performed through adjectives as "Bad", "difficult", "favorite", "fine", "good", lovely, etc. Another representation quite common is that of superiority, a more pronounced type of the evaluative carried out through adjectives like "best", "greatest", "wisest", etc. Physical representations were also very frequent ("attractive", "beautiful", "gallant", "handsome", etc.). The age grading also proved to be a present representation ("earliest", "Early", "later", "old", "older", etc.). Clinical conditions have also been shown to frequent representation ("autistic", "crazy", "deaf", "depressed", "ill", etc.) as well as types of behavior ("aggressive", "angry", "bored", etc.), social issues (e.g. "Battered", "disadvantaged", "homeless", "unemployed", etc.), inferiority (e.g. "Desperate", "inexperienced", "lesser", "primitive", etc.) and gender identification (e.g. "Bisexual", "heterosexual", "homosexual", etc.). This set of representations shows the main characteristics attributed to the human being from the point of view historic. As far as we can verify, this is the first description of this type in the literature. The analysis detailed the representations of each term and showed two results unexpected: what seemingly synonymous terms have distinct representations and that morphological forms of the same term have separate representations. Regarding the first point, the analysis showed that the various terms referring to the adolescence (adolescent/s, teen/s and teenager/s) bring a mix of representations different hues: "adolescent/s" place greater emphasis on gender identification; teen/s, mental aspects; and teenager/s, to marital status. But they all have an emphasis main point, as stated, the evaluative, coloring not very favorably, that phase of life. Regarding the second point, the analysis showed differences between representations around the singular and plural forms of the term. For example, gender identification is most conveyed through the form plural than the "adolescent" singular. This confirms once again the finding of Corpus Linguistics that the language in use avoids true synonyms, as each form tends to assume a different role. As a possible reply to the second research question, the analysis showed that there are marked differences between representations from both a historical point of view and age and gender. From a historical point of view, there has been an increase identified in the representations of stages of life recognized as "elastic" categories, which can be graduated by means of adjectives like "young", "old", "senior" and even "adolescent". Furthermore, in comparison to the beginning of the 19th century, there was an increase in representations of gender identification feature that goes beyond the binary classification, through "Homosexual", "lesbian", "bisexual" and related adjectives. There was also an increase in ethnic-racial representations, such as "black" and a decrease in others as "colored". From the age point of view, marked differences were identified: childhood, is marked by physical representations (barefoot, beardless, beautiful), of behavior (aggressive, happy, idle, merry, etc.), age (oldest, preschool, senior, etc.) and superiority/virtue (popular, promising, smart, etc.). The adolescence is marked by behavioral representations (angry, bored, noisy, etc.), social (disadvantaged, runaway, unemployed, etc.), clinical (deaf, depressed, disabled, etc.), gender identification (bisexual, lesbian, etc.) and inferiority (impressionable, inexperienced, etc.). In contrast to these groups, adult life is represented in a very marked way by notions of superiority/success (ambitious, civilized, distinguished, eminent, etc.), occupation (literary, medical, military, etc.) and inferiority (desperate, lesser, primitive, etc.). In summary, the conception of human being varies throughout life, passing from a figure classified by their physical appearance, age nuances and potential, to a generally conflicted figure, marked by medical issues, of identification and behavior, ending in a figure again classified by his physical attributes, but seen through their occupation and success. Finally, in relation to the third research question, the analysis showed great variation between terms. Regarding gender, the results suggest that the terms generic male (man and men) tend to have a more positive valuation than all female correspondents do do. At the same time, feminine terms are more positively represented than the males who designate youth (boy and boys). Regarding age progression, the analysis suggests, in general, that the adult life is more positively represented than childhood, which in turn is more positively represented than adolescence. However, the disparity between the terms of each group: between the terms "adults", "Man" is the best rated, but "adults" is four positions (15 out of 20) from the last ranking position. The same occurred with the other groups, to a greater or lesser extent degree. In all cases, however, the best-valued item tended to be singular, while the worst valued, in the plural. The research presented here has limitations, such as the impossibility of deal directly with texts from Google Books. The data used, provided by Google, are restricted to lists of bigrams, which include the year of publication and the number of occurrences that year. The text itself is not available. Per this, it was not possible to raise the textual occurrences and make the analysis from them. This limited the analysis to bigrams, instead of the entire text of the works. Another limitation is that some terms are not exclusively age or gender indicators, as "man", which is used to identify the human being in general (male and female), and "boy" and "girl", which can be used to refer to people of various ages, figuratively. The survey offered, for the first time, an overview of how life is historically represented in the English language, with respect to its phases and differentiation between genders. The research showed the impossibility of broad generalizations: each term has its own range of representations, which distinguish it from other terms, even those closest conceptually or morphologically. The language in use resists broad generalizations: there are many nuances between terms. The generalization we can make, based on the results, is that the passage of life is marked by a constant classification of the human being in terms of a finite set of representations, with notably evaluative and normative bias. Moreover, that passage is marked historically, with marked temporal changes in the last 200 years.

Acknowledgements

I am very grateful to the Coordination for the Improvement of Higher Education Personnel (CAPES) for the scholarship granted, since the first semester of the doctorate course, and that provided me the chance of realization of a Doctorate course. Process # 773985 (Portal PUCSP, São Paulo Catholic University in Brazil Southern State) and last, but never least, Professor Tony, Berber Sardinha for being my advisor throughout this research study. Also, I am thankful to my family who has always supported me.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Baker, P. (2014). *Using Corpora to Analyze Gender.* Bloomsbury.

Baker, P., & Ellece, S. (2011). *Key Terms in Discourse Analysis.* Continuum.

Baker, P., & Potts, A. (2013). Why Do White People Have Thin Lips? Google and the Perpetuation of Stereotypes via Auto-Complete Search Forms. *Critical Discourse Studies, 10,* 187-204. https://doi.org/10.1080/17405904.2012.744320

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016a). *Diachronic Word Embeddings Reveal St.*

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016b). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* Berlin, August 2016, 1489-1501. https://doi.org/10.18653/v1/P16-1141

Stubbs, M. (1996). *Text and Corpus Analysis Computer-Assisted Studies of Language and Culture* (p. 98). Blackwell Publishers.