A Proposal for Genre Analysis: The AMS model

Ricardo Casañ-Pitarch
Universitat Jaume I, Castellón, Spain
casanr@uji.es

Abstract

Genres tend to be studied from analytical and professional perspectives. However, the study of genres can also be used for pedagogical aims in language teaching (Bhatia, 2002). Writing for academic and professional purposes as well as for scholarly publication is often challenging. Some novice writers may find difficulties when they work on new genres and fail to recognize the communicative purpose, organization of the texts and linguistic rules. This paper focuses on a self-developed model of analysis of multi-genre structures (AMS) which aims at determining the common features within a corpus of documents of the same nature. The objective of this AMS model is to help researchers determine the form of any genre related to the users’ personal or professional fields; this involves the study of macro- and micro-structures. This paper begins with a general review of genre analysis theories and connects with the use of suitable corpus linguistics software. Then, our model of genre analysis to carry out is introduced. At last, results on previous analyses are presented with the aim of explaining how to apply the AMS model and discuss their use.

1 Introduction

The evolution of language along history is an unquestionable fact. Noticing changes in one’s own language is as simple as reading a book published some hundred years ago; for example, British and Spaniards would find remarkable differences after reading a few lines written by Shakespeare or Cervantes and then comparing them with contemporary literature in British English and Continental Spanish respectively. There are different arguments that may justify the evolution of language; however, it seems that one of the most reasonable facts for such evolution is the changing communication needs in society. Language is a social tool created and used by individuals to communicate; thus language evolution is the result of changes and new inventions in society that make necessary that utterers from a particular community adapt their discourse and create new linguistic forms in order to communicate with other individuals, conveying their messages successfully as well as making them understandable for their receivers. In this sense, it can be assumed that language and society have evolved along history hand-by-hand; however, it seems that social changes are the main reason for language evolution and not the other way round. With this premise that society modifies language, it seems to be logical that
this fact influences on how fast languages changes. The evolution of society during the XX century was huge and there is clear evidence that this will not be different during the XXI, but changes will probably be more noticeable. The creation of the Internet and making it accessible to most homes as well as the development and spread of media throughout the globe have made that information can be transferred faster and the amount has vastly increased. This increase of the information transferred has affected to all the communication fields such as business, medicine, education, politics, or general and daily situations. In this context of continuous changes in language, linguists play an important role to describe and explain the new forms of language and its use. This paper is a response to these social needs regarding language use and focuses on introducing a model of analysis of multi-genre structures (AMS). The function of this model is to analyze a corpus of texts with the same particular nature and define their standard or prototype form following statistical analyses. Once determined the standard forms of a particular genre, these results could be used to define or help reproduce the target genre for professional, academic, or pedagogic purposes.

2 Theoretical Principles of Genre Analysis

Genres can be defined as “a class of communicative events that share some set of communicative purposes” (Swales, 1990). In order to describe these communicative purposes, it is necessary to focus on their content, form, and functionality (Shaw, 2014), and this analysis process is known as genre analysis. As previously explained, genre analysis is a social communication need (Bhatia, 1993; Dudley-Evans, 1986; Swales, 1990), which seems to have gradually increased due to the demands of communication derived from the effects of globalization and development of media. The analysis of language enables researchers to identify patterns of forms and function in particular genres and areas of language. Bhatia (2004) defined genre analysis as a multi-disciplinary activity used to determine the form of any particular type of text and to help individuals reproduce them with the correct choice of lexicon and grammar. In this sense, new and hybrid genres are constantly appearing in form of emails, commercial brochures, websites, blogs or in any other written or oral channel. It shall be acknowledged that genres can share both similarities and differences; thus, each genre contains specific features that require the use of a particular appropriate language code (Pfenniger, 2008). This variety of generic social discourse needs to achieve specific communicative purposes within different contexts; and these become more specific with the involvement of individuals’ private intentions within the contexts of socially recognized communicative purposes (Fairclough, 2013). These facts make that language forms become unique for each of the existing varieties and sub-varieties of any existing genre.

In order to carry out an analysis of genres, there are different items that need to be considered with the aim of determining the standard form of the target genre. Some authors have provided their own models of genre analyses (Bhatia, 1993; Hyland, 2000; Killingsworth & Gilbertson, 1992; Lassen, 2003; Swales, 1990); some of them were original and some others were based on previous models. However, all of them aim at providing a versatile and dynamic tool with an explanatory vision, which allows for continuous innovation and exploitation of documents (Bhatia, 2004). One of the most interesting models regarding our research interest based on morphology and syntax levels is Bhatia (1993). As it can be observed in table 1, this model consists of seven stages, and only the number six refers to language analysis. The other six items are relevant for contextual settings, but these do not directly interfere to the language forms that each genre is characterized by.
As explained in Casan-Pitarch and Calvo-Ferrer (2015:77), the sixth stage in the model of Bhatia (1993), ‘linguistic analysis’, focuses on the analysis of the most significant features of language and this involves three levels:

- **Level 1: Lexicon-Grammatical Features.** This analysis provides empirical evidence to confirm or disprove the use of certain words in the target genre. This level aims at studying the grammatical features and any specific features of language common to the genre. This analysis is statistical and it requires a relatively large corpus. This type of analysis is limited since it cannot explain the use of certain words, and how these are involved in the process of communication.

- **Level 2: Text patterning.** The second level focuses on the language strategies, which specify how the members of a community give specific values to the language use. This may involve the study of lexicon, syntax or discourse in any particular genre. In other words, this level describes the choice of certain language forms.

- **Level 3: Structural interpretation of the text genre.** The last level focuses on the analysis of the structure and organization of the genre regarding the study of moves and steps.

### 3 Description of the AMS Model

Based on Bhatia’s (1993) previous model of genre analysis at the linguistic level, a model for the analysis of multi-genre structures has been developed. The purpose of this AMS model is to analyze, describe and reproduce genres through analytical research of either oral or written discourse forms (see figure 1). The proposed model is divided into two parts as it is shown in table 2: macro-structure and micro-structure (Casan-Pitarch, 2015a).

<table>
<thead>
<tr>
<th>Macrostructure</th>
<th>Quantification of forms and structures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moves and steps forming the genre</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Microstructure</th>
<th>Language Typology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Morphology</td>
</tr>
<tr>
<td></td>
<td>Morphological Formality</td>
</tr>
<tr>
<td></td>
<td>Terminology</td>
</tr>
<tr>
<td></td>
<td>Verbal analysis</td>
</tr>
<tr>
<td></td>
<td>Analysis of personal pronouns</td>
</tr>
<tr>
<td></td>
<td>Syntax</td>
</tr>
</tbody>
</table>

**Table 2:** Model for the analysis of multi-genre structures (AMS) at Linguistic level

### 3.1 Macrostructure

The macrostructure of a genre refers to the study of the major structures and forms. Van Dijk (1995) explained that macrostructures focuses on semantic structures that describe the overall unity and coherence of the text from a global perspective. Ruiz-Moneva (2011) added that macrostructures concerns the meaning of the text. As it can be observed in table 1, analyzing the macrostructure involves the quantification of forms and the description of the moves and steps in the text.
3.1.1. Quantification of Forms and Structures

The first stage in the AMS model is the quantification of forms and structures. This analysis is used to determine the amount of words, sentences, and paragraphs. With this data, there is different textual information that can be registered such as words per sentence, sentences per paragraph and paragraphs per text. The analysis of the macrostructure is fundamental to set the basis and the main forms of the target genre. As a result, the general structure of the target genre can be visualized before analyzing the different elements of the microstructure in detail. In a metaphoric language, this part of the research would represent building the skeleton of a house. The collection of this data can be achieved with the software ‘Wordsmith’; this tool enables individuals to automatically quantify words, sentences, and paragraphs. The following Figure shows an example of a corpus which has been automatically analyzed with ‘Wordsmith’ (6.0) and the results obtained, which include the amount of words, sentences, paragraphs, or headlines among others.

![Figure 1: Wordsmith tools (6.0) for quantifying forms and structures](image)

3.1.2. Moves and Steps Forming the Genre

Once the forms of the text have been quantified, the second item in the analysis of the macrostructure concerns the identification of moves and steps within the genre. Swales (1990) defined moves as parts of the text with particular communicative purposes, and these are divided into steps. Thus, steps are the division of information which aims at realizing the communicative purposes indicated by the move (Bhatia, 1993). In other words, moves could be understood as the purpose of a part of a text, whereas steps would be the contents or information used to fulfill that purpose. In this case, electronic analysis does not seem to be possible; thus, manual analysis is required to obtain the necessary information at this stage. Returning to the language of metaphors, the results obtained at this level shall be used to continue building the house whose skeleton had previously been made with the quantification of forms and structures; in this case, the walls of the house and the division of rooms will be made.
1. Establishing a Territory

- Step 1. Establishing centrality
- Step 2. Making topic generalization
- Step 3. Reviewing Items of previous research

2. Establishing a Niche

- Step 1a. Counter claiming, or
- Step 1b. Indicating gap, or
- Step 1c. Question raising

3. Occupying the Niche

- Step 1. Outlining purposes
- Step 2. Announcing present research
- Step 3. Indicating research article structure

Table 3: Example of the moves and steps available in the analysis of introductions (Swales, 1990)

3.2 Microstructure

While macrostructures concern the study of the major forms, microstructures should refer to minor forms within the target genre. Glowalla and Colonius (1982) explained that the microstructure corresponds to the study of individual sentences and the connection among sentences and words. In this sense, our model includes the analysis of a series of items at morphology and syntax levels: type of language, morphology, terminology, formality and verbal tenses. It shall be acknowledged that this list can be extended to other items depending on the researchers’ needs, interest, knowledge and experience. In the language of metaphors, the microstructure of the texts would correspond to modeling the interior of the house (kitchen, bathrooms, floors and walls, etc).

3.2.1. Language Typology

The first item in the analysis of the microstructure focuses on the study of the type of language. In this sense, the language can be classified into different categories; the major text types are narrative, descriptive, directive, expository, and argumentative (Schneider, & Barron, 2014). In this case, the software ‘Tropes’ (8.44) offers some clues after an electronic analysis of the lexicon. Figure 2 shows an example of this electronic tool.

![Figure 2: Example of language typology analysis completed with ‘Tropes’ (8.44)](image)

On the other hand, there is another possibility to classify language into informative, persuasive, and promotional varieties (O’Hair et al., 2004). This classification is simpler than the first and the classification could be considered easier. The aim of informative language is to increase the audience’s understanding or awareness by offering knowledge (Soltero, 2011). O’Hair et al. (2004: 349) described persuasive language as a creation to evoke reactions in listeners, which is meant “to appeal to the audience’s attitudes, beliefs, and values about the issue in question and to sway listeners to the speaker..."
point of view”. The language of promotion focuses on marketing interests, and it is a way of being persuasive without looking like that (Bhatia, 2005; Jendrych, 2011). This stage requires manual analysis, and this purpose can be achieved by classifying sentences, paragraphs or the whole document into types of language. The collection of this information shall be used as the basis to determine the most common type of language in each step.

3.2.2. Morphology

The second analysis of the microstructure concerns the morphological composition of the genre. Morphology is defined as the words in a language (Shore, Ray & Goolkasian, 2013). This analysis involves the quantification of the different word categories into percentages: nouns, determiners, adjectives, prepositions, verbs, pronouns, adverbs, conjunctions and interjections. With this analysis the most common word composition of a genre can be explained. It is a fact that the use of certain word categories is more common than others (e.g.: higher nominalization vs. low use of verbs; higher or low use of adjectives); thus it seems that this analysis is necessary in order to explain the form of a genre. This type of analysis can be partially completed with the software ‘Tropes’ (8.44). It quantifies most different word categories (nouns, verbs, adjectives, conjunctions, adverbs and pronouns) except determiners, prepositions and interjections. This lack of data can be overcome with a manual check by searching the determiners (the, a, some, any, this, that, these, those, etc), interjections and prepositions in the text. See figure 3 showing a view of the tool for morphology analysis with ‘Tropes’.

![Figure 3: Example of morphology analysis with ‘Tropes’ (8.44)](image)

3.2.3. Morphological Formality

The following item in the AMS model is the analysis of morphological formality. As it has been previously commented, genres and texts contain different types of word categories. Using the data obtained in the previous stage, morphology, Heylighen and Dewaele (1999) suggest that the level of morphological formality can be obtained with the formula shown in table 4. They state that words are divided into two broad categories. Words can be context independent (formal) and context dependent or deictic (less formal). Formal categories are nouns, adjectives, prepositions, and determiners; whereas pronouns, verbs, adverbs, and interjections are deictic forms. Conjunctions are considered neutral. This process cannot be completed electronically, but it is very simple if the user has obtained the necessary data regarding morphological categories.
3.2.4. Terminology

The analysis of the terminology within the corpus of texts offers valuable data about the most frequent words. These results are useful to show the relevance of certain words in the text, and consequently their presence should be more or less obligatory. The software ‘Wordsmith’ can analyze the most common words with no distinction of their morphological category; however, the software ‘Tropes’ can identify the most usual terminology and form groups of words according to their morphological categories and divide them into lexical groups such as business, politics, locations, people, times and dates, health, arts, among some others. In addition, these categories are also divided into subcategories; for example the area of business would include accounting, banks, brokerage, economy, money, or insurances. This classification of words is useful to determine the type of words that are needed to develop the standard form of any particular genre. Figure 4 shows an example of the analysis of terminology with ‘Tropes’ (8.44).

3.2.5. Verbal Analysis

The next item in our list concerns the verbal composition of the genre. This classification is based on Downing (2014) and focuses on determining the finiteness, tense, anteriority, mood, emphasis, aspect and polarity of the verbs. Verbs are a class of linguistic signals that are used in sentences and convey actions, or expresses states or relation among things and/or people (Cangelosi and Parisi, 2001). In communication, verbs provide the focal point of the clause and control the other elements occurring in the clause (Biber, Johansson, Leech, Conrad, & Finegan, 1999). This analysis offers valuable information concerning the type of actions in the text contextualizing time and style. This analysis cannot be carried out electronically; at least we have not identified any software that permits doing so. The following table can be used to list verbs following Downing’s (2014) classification.
3.2.6. Analysis of Personal Pronouns

Pronouns are defined as words which are used in place of nouns and noun phrases and they usually refer to element referring to people, animals, ideas, or things (Schmidhauser, 2010). The main function of pronouns is to shorten sentences and make communication easier (Manning, 2005). As explained in Casan-Pitarch (2016), people modify the form of their messages according to their interests and purposes. Thus, linguists can do research on linguistic norms that determine why a pronoun is used instead of a noun or another pronoun. In this sense, this analysis focuses on the use of personal pronouns and it also considers possessive determiners due to their effects of personalization in the sentence. The analysis of personal pronoun accepts electronic support, but they actually need supervision. Personal pronouns are limited in number: subject, object, possessive, reflexive pronouns and possessive adjective pronouns. For this reason, researchers can find them electronically in the text and quantify the occurrences. The only trivial aspect that requires supervision is the case of pronouns whose function is variable. In this case, the pronouns ‘you’ and ‘it’ are used as subjects and objects, ‘her’ can be either object or possessive pronoun, whereas ‘its’ and ‘his’ can be possessive adjectives or pronouns at the same time. In these cases, it is necessary to check the form the pronouns refer to.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>I</td>
<td>Me</td>
<td>My</td>
<td>Mine</td>
<td>Myself</td>
</tr>
<tr>
<td>Plural</td>
<td>We</td>
<td>Us</td>
<td>Our</td>
<td>Ours</td>
<td>Ourselves</td>
</tr>
<tr>
<td>Second</td>
<td>Singular/Plural</td>
<td>You</td>
<td>You</td>
<td>Yours</td>
<td>Yourself</td>
</tr>
<tr>
<td>Third</td>
<td>Feminine</td>
<td>She</td>
<td>Her</td>
<td>Hers</td>
<td>Herself</td>
</tr>
<tr>
<td>Masculine</td>
<td>He</td>
<td>Him</td>
<td>His</td>
<td>His</td>
<td>Himself</td>
</tr>
<tr>
<td>Neutral</td>
<td>It</td>
<td>It</td>
<td>Its</td>
<td>Its</td>
<td>Itself</td>
</tr>
<tr>
<td>Plural</td>
<td>They</td>
<td>Them</td>
<td>Their</td>
<td>Theirs</td>
<td>Themselves</td>
</tr>
</tbody>
</table>

Table 6: Analysis of Personal Pronouns

3.2.7. Syntax

Harley and Noyer (2000:3) defined syntax as “the constituent structure of strings of words”, while Fogassi and Ferrari (2007:140) referred to syntax as “a rule-based system combining elements into a sequence that has a specific meaning”. These definitions suggest that syntax is a science that studies the connection and meaning of words in sentences and clauses. The last stage in the AMS model concerns the analysis of syntax. At this stage, the corpus needs to be analyzed individually and different aspects can be considered. The first analysis focuses on the classification of the clauses into simple, compound, or compound complex. Then, the study continues with an estimation of the average extension of the subject and predicates. With these results, the average global structure of the sentences can be defined. In addition to these parameters, this study also focuses on the quantification of the most common functional items within the sentences. On the one hand, this model suggests quantifying the functional composition of the sentences (direct objects, subject complements, prepositional objects, adjunct or conjuncts, etc.). In addition, this analysis also considers the study of the most common functional structures within the sentences (e.g.: ‘S+V+DO’, ‘S+V+C’, ‘S+V+SC’). Regarding our experience in this type of analysis, it is difficult to determine a standard model of syntax because it tends to be quite varied; however, these results are a valuable resource of information that can be used as reference.
4 Data Collection and Procedure

In order to use the AMS model, it requires specific instructions for the collection of data and also information about the procedure to obtain valuable results. The application of this research model requires different types of analysis as well as the interpretation of results with the aim of creating some useful guidelines that may determine the standard form of the target genre. The most usual way of analysis is the calculation of the mean, median, mode, range and usual range. These terms should be individually defined in order to understand this analytical process. To begin with, the mean is the average of all the items analyzed or considered (Brown & Kim, 1991). The median is defined as “the value having half of the observations less than and half exceeding it” (Campbell & Gardner, 1988:1454). The mode is the most frequent value within the data; and this can be more than one value (Larson, 2006). In addition to these definitions, the range and usual range need to be considered. On the one hand, the range is the data or area between the lowest and the highest levels (Hanley, 1993). On the other hand, the usual range is defined as the most common continuous levels of frequency (Kivilis & Gafanovich, 1964). This variable is the most difficult to establish because it depends on the other four. Whereas the mean, median, mode and range are independent factors as well as precise and relatively easy to calculate, the usual range corresponds to the series of values which are considered the most common or the prototype form. If the levels of the mean, median and mode are similar, the usual range will be precise and clear; this should be considered a valid usual range. However, it also may happen that these values differ largely and the usual range cannot be established; this should be considered an invalid usual range. Despite this limitation, the researcher may try to find an explanation to any invalid usual range. In fact, the existence of two or more prototype forms is also possible; in this case, the corpus could be divided into blocks sharing certain similarities and then establish their variables with the aim of obtaining their respective standard forms. As it can be observed in Figure 5, the usual range is valid when the mean, median, and mode contain reasonably similar values; the usual range is also a method to shorten the range. In the example below, the range involves eight levels whereas the usual range only two.

![Figure 5: Example of data procedure](image)

Despite this previous model of data collection and procedure is the most common within the AMS model, another way of analysis is used. In this sense, this previous analysis is only applied to the macrostructure and the morphology in the microstructure. The method of data procedure for the other elements in the microstructure is based on the best results. In this sense different options are possible, but the guidelines established according to the results obtained for the development of the standard model suggests the most common forms without being restrictive to other possibilities. For example, regarding the analysis of verbs, the guidelines may suggest the use of the active present simple, but it cannot ban the use of other forms since this may vary according to individual needs and purposes.

At last, after placing together the results of the macrostructure and the microstructure, the standard model of the genre can be reproduced. We suggest proving the validity of the results obtained and the parameters established in the guidelines with the creation of a sample of the target genre and then getting...
a specialist to check the text (Bhatia, 1993). This reproduction should include all the steps and moves that are considered compulsory and optional according to the criteria established, and the language used should follow the standard parameters obtained from the analysis of the results. This analysis can also be complemented with a particular format. This information may be obtained with an additional analysis focusing on elements related to appearance of the document such as the font type, font size, font color, line spacing, text alignment, background color, word emphasis, videos, images, sounds and links, among others (Casan-Pitarch, 2015b).

5 Conclusion

Analyzing genres and establishing specific guidelines explaining their standard forms or the most common features is the result of the present communication needs in society. As it has been explained, it is unquestionable that communication has largely increased in the last decades, especially once the Internet was established at most homes and more recently to portable devices such as smartphones and tablets. Thus, it seems that the analysis of genres has become a major social need in order to communicate well with the correct forms in any contemporary context. The aim of this paper has been to suggest a model of genre analysis and to explain its use and how it works. This model of genre analysis is addressed to communication professionals from different fields and it provides them with a tool that helps analyze the most specific texts and determines genres based on their professional or academic needs. The AMS model includes the analysis of different parameters that may be used to describe and reproduce genres through analytical research of either oral or written discourse forms. This model can be used either for professional or educative purposes, and the final aim is to develop materials and documents adapted to specific contexts and needs. As it has been commented, the AMS model is divided into two stages: macrostructure and microstructure. Each of these analyses contains different items. Some of these can be analyzed electronically, whereas others need manual analysis. This fact proves there is a need to further research in the area of corpus linguistics in order to facilitate the processes to analyze genres.

References


Shore, R., Ray, J., & Goolkasian, P. (2013). Too close for (brain) comfort: improving science vocabulary learning in the middle grades: researchers take a multidisciplinary approach to

