

# A Linguistic Analysis of Startups in The Context of the Air Transport Industry Management

Olga Zervina<sup>1</sup>[0000-1111-2222-3333]

<sup>1</sup> Transport and Telecommunication Institute, 1 Lomonosova street, LV-1019, Riga, Latvia  
zervina.o@tsi.lv

**Abstract.** Much research has studied how a company can maximize its profit. Relatively small number of them focused on Value Proposition, though the number of authors proved that companies that deliver multiple values experience better business performance. This article describes a current research on linguistic analysis of startups in the context of the air transport industry. Analyzing startups manually is a very time consuming task, so the automation of the process would be beneficial. The author takes corpus linguistic approach, created an experiment protocol and is on the stage of conducting an experiment. Under this experiment air transportation startups' landing pages were collected in the number of 800. 100 annotators first were preliminary surveyed and then trained to annotate startups. Post-annotation training will be conducted to understand the difference in expertise level. The annotation results will be further analyzed and linguistic features and patterns will be identified. As a result of the research, an author will develop a methodology for analysis of values based on a model of automatic identification of values in the text of a startup's landing page in the air transportation industry.

**Keywords:** Startups, Value Proposition, Air Transport, Annotation.

## 1 Introduction

Value chain analysis is a primary corporate strategy tool. Traditionally relatively few values have been pursued; price, quality, etc. Recently, customers are demanding a wide variety of values. Most importantly, some of the values that have previously received little attention become dominant values (concept known as value shift). Some examples: cars - eco-friendliness, electronics – usability, food - fairness, organic. As new values become prominent, industries and their companies will be transformed; not following values could result in wasted resources (pursuing the wrong values) and becoming irrelevant (not pursuing the right ones). Hence, the ability to identify values early on is important for business performance. Startups are often the first to discover new values. As an example, *Skytran*, startup from air transportation field, offers autonomous, zero-emission vehicles arrowing above congested streets.

The company’s landing page identifies its values as “high speed, high capacity, low cost” [1]. Analyzing startups manually is a very time consuming task. The author assumes that one startup landing page takes 10 minutes to examine. Air transport industry offers around 10000 startups per year, requiring 1 full-time analysts to keep track of startups of just one industry on annual basis; and there are 100’s of intersecting industries in aviation (food & beverage, hospitality, etc.). Methods that allow at least in part automate this process would be beneficial. This research describes the first stages of developing a methodology of analyzing values within industry based on startups value propositions.

**The main research question is** tools and procedures for analysis of value proposition of startups in the area of air transportation from their textual description in the frame of developing a methodology for analysis of values based on a model of automatic identification of values in the text of a startup’s landing page in the air transportation industry.

**The initial contribution of this research** is the development of methodology of analyzing values within industry based on startups value propositions. The procedures and tools integrated in the model are either articulated or developed by the author throughout the Thesis.

**The approach.** The author uses a Natural Language Processing (NLP) approach to identify the methods and features well suited for this problem. A bottom-up (data-driven) technique is taken, i.e. the author first constructs the dataset and then analyzes it using both computational and linguistic methods to identify which features and methods perform the best. The author also takes the corpus linguistic approach, i.e. reliable language analysis is more feasible with corpora collected in the field in its natural context.

**The object** of the research is startups’ landing pages in the field of air transport industry.

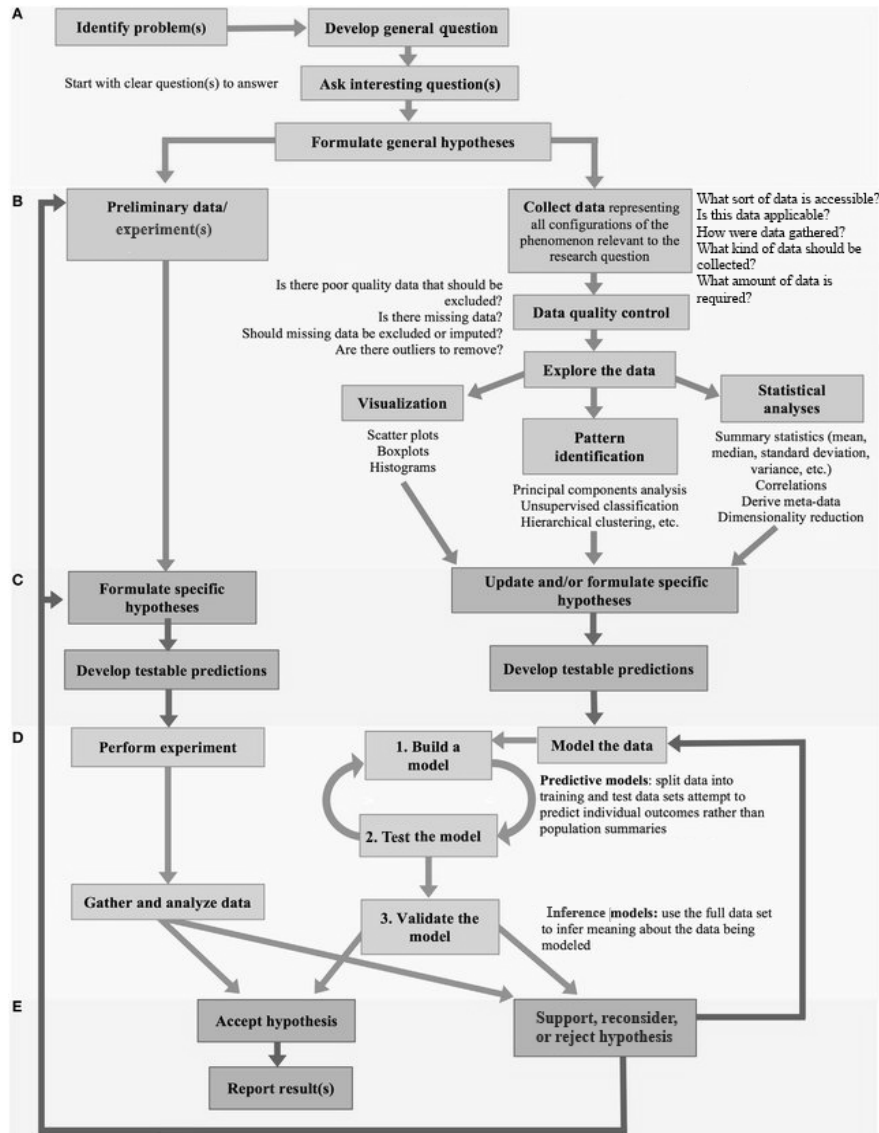
**The subject** of the research is value proposition of a startup.

**The tasks** to be performed are the following:

1. To study theoretical literature on the research topic
2. To examine existing research on the research topic
3. To collect preliminary data
4. To conduct an experiment
5. To identify features (e.g. linguistic, semantic)
6. To identify patterns
7. To develop testable predictions
8. To build a model
9. To test a model

## 2 The Methodology and Methods of Research

The research methodology is presented in Figure 1. The author takes *data-driven scientific method*. The general idea was adopted from the paper describing the scope of big data in medicine [2].



**Fig. 1.** Stages: (A) Framing the problem and general hypotheses. (B). Data collection and exploratory experimentation/analysis. (C) Formulation of specific hypotheses. (D) Testing the hypotheses. (E) Accepting or rejecting the hypotheses.

The research hypotheses were formulated as follows:

*general hypothesis:* values could be automatically identified from text (landing page of a startup)

*specific hypotheses:* 1. Significant increase in number of known values is possible even for people with higher level of domain familiarity; 2. Linguistic features can improve accuracy of value identification.

With the accessibility of large datasets and advanced correlation/statistical tools, will we still need to rely on hypotheses in scientific research? Traditionalists say that in purely data-driven methods, one may not realize where to look for those riveting findings if no hypotheses were formed beforehand.

Big data advocates, on the other side, propose that with no prior beliefs, one is not driven by established ways of thinking or creating, opening the possibilities of breakthrough insights where nobody had been before [3].

**Research Theoretical Framework.** The following theories and methods have been used in work: the system approach [4], information extraction, exploratory experimentation [5], the methods of statistical analysis [6], simulation modelling approach [7], methods of quality analysis [8], surveys, pattern identification [9], building a model.

**Limitations of the research** are linked to the research base: one industry for startups - aviation industry, one hundred experiment participants and their level of expertise (undergraduate students, non-native English speakers), task complexity and ambiguity, ambiguity of natural language.

**Scientific novelty:**

analysis of Value Proposition in the context of air transportation and startups; construction of Value Proposition corpus / dataset; systematization of Value Proposition (ontology); analysis of linguistic patterns of Value Proposition; automation of linguistic features detection that could be used to detect Value Proposition (feature engineering [10])

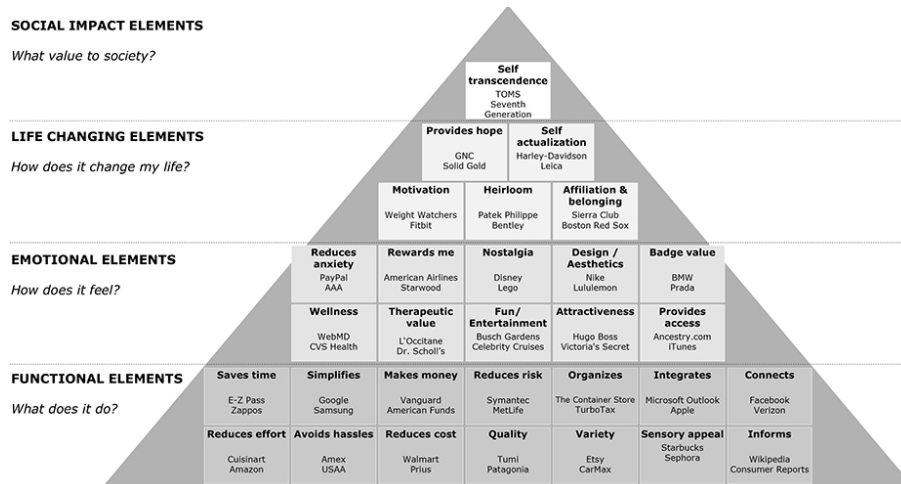
### 3 State of the Art

Articles reviewed here and referenced throughout the article were sourced from online libraries using a relative research approach: papers related to Value Proposition, startup concept, corpus linguistics, text annotation. The inclusion criteria for an article in this review were the following: the study was of an exploratory or empirical nature or gives an idea of the terminology source.

A rich theoretical and empirical literature can be applied to the question on how to identify Value Proposition in startups texts. As the current research focuses on linguistic patterns of startups landing pages, the author reviews theoretical and empirical literature on specific online text features, textual value proposition methods, value proposition comprehension by consumers of different level of expertise.

Michael Lanning and Edward Michaels first used the expression “value proposition” (VP) in a 1988 work document for the consulting company “McKinsey and Co”. In the article, which was entitled “Delivering value to customers”, the authors define value proposition as “a clear, simple statement of the benefits, both tangible and intangible, that the company will provide, along with the approximate price it will charge each customer segment for those benefits”[11].

In 2016, Eric Almquist suggested a strategy based on a differentiated customer value proposition. A suggested set of value was called Elements of Value [12]. Elements of Value categories are based on Maslow’s Hierarchy of Needs shown in Figure 2.



**Fig. 2.** Heuristic model of value with examples of companies exhibiting elements of value [12]

There are comparatively few researches on analyzing value propositions in online startups. In 2007, Su-C Li in his paper argues that a properly constructed value proposition is essential to the value creation process in e-business, and value co-production is the building blocks for value protection mechanism in network economy [13].

Äyväri, Anne, and Annukka Jyrämä in their article “Rethinking value proposition tools for living labs” published in 2017 provide a conceptual analysis on value proposition tools to be used in future empirical research and in building managerial insight. The conceptual analysis focuses on a living lab framework and recent theoretical developments around the concept of value that are reflected in the context of three managerial tools for creating value propositions. Among findings in the context of the living labs approach, the Value Proposition Builder seems to conflict with the ideas and premises of user-centric innovation processes [14]. In 2019, Guo, Hai, Jun Yang, and Jiaping examined the fit between value proposition innovation and technological innovation (exploitative vs explorative) for the performance of startups in the digital environment. They based their research on on-site survey data of 285 digital startups in one of the world's largest digital economies and found that explorative innovation strengthens the positive impact of value proposition innovation on the performance of startups, but exploitative innovation weakens this positive effect [15].

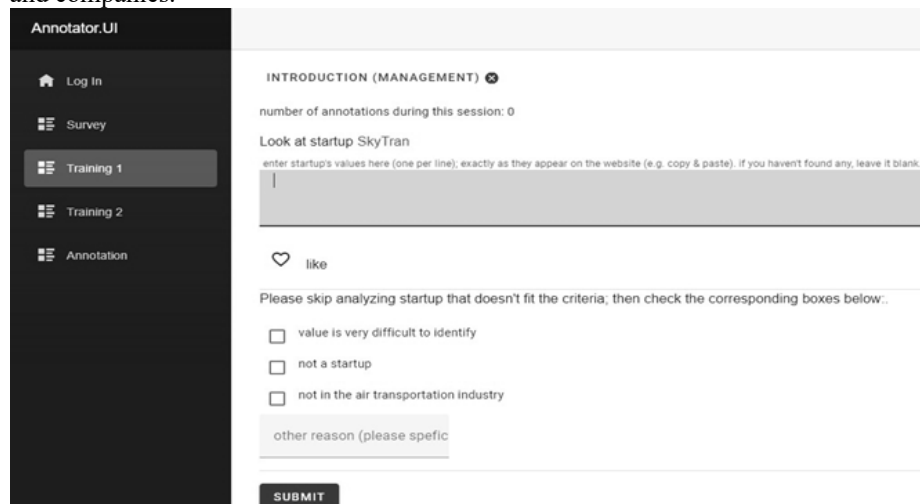
Corpus linguistics has generated a number of research methods, which attempt to trace a path from data to theory. Wallis and Nelson in 2001 first introduced what they called the 3A perspective - Annotation, Abstraction and Analysis: Annotation consists of the application of a scheme to texts [16]. Annotations include structural markup, part-of-speech tagging, parsing, and numerous other representations. Abstraction consists of the translation (mapping) of terms in the scheme to terms in a theoretically motivated model or dataset. Abstraction typically includes linguist-directed search but may include e.g., rule-learning for parsers. Analysis consists of statistically probing, manipulating and generalizing from the dataset. Analysis might include statistical evaluations, optimization of rule-bases or knowledge discovery methods.

## 4 Research Design and Preliminary Results

### 4.1 First Stage – Current

A total of 800 different startups in the field of air transportation were chosen, their landing pages consist the research base for analysis on Value Proposition. 100 participants of low level expertise - IT, Aviation, and Management undergraduate students – were involved in a preliminary survey, post-survey, training and post-training annotation of startups landing pages. A separate webpage was created for conducting surveys, two-level training and annotation process. The quality of annotation as it was performed by low-expert students was assessed hereafter by industry experts.

The aim of a preliminary survey is to understand the level of expertise for non-trained participants. The students were told about the basic concept of value proposition. Also, they were given some examples of Value Proposition such as typical ones: affordability, quality, speed; less typical: eco-friendliness. After that they were asked to list as many values as they can, that are provided by the Air Transportation startups and companies.



The screenshot shows the 'Annotator.UI' web application interface. On the left is a dark sidebar with navigation options: 'Log In', 'Survey', 'Training 1' (highlighted), 'Training 2', and 'Annotation'. The main content area is titled 'INTRODUCTION (MANAGEMENT)' and shows 'number of annotations during this session: 0'. Below this, it says 'Look at startup SkyTran' and provides instructions: 'enter startup's values here (one per line), exactly as they appear on the website (e.g. copy & paste). if you haven't found any, leave it blank.' There is a large text input field. Below the input field is a 'like' button with a heart icon. A note says 'Please skip analyzing startup that doesn't fit the criteria; then check the corresponding boxes below..'. There are three checkboxes: 'value is very difficult to identify', 'not a startup', and 'not in the air transportation industry'. Below these is a text input field for 'other reason (please specific)'. At the bottom is a 'SUBMIT' button.

**Fig. 3.** The experiment webpage with training 1 procedure Source: <http://ttiv.s3-website-us-east-1.amazonaws.com/>

Figure 3 shows the experiment webpage structure, where participants conduct a survey, two trainings and an annotation procedure. The Figure reflects the Training 1 procedure. At the end of annotation experiment the participants are asked to conduct a new survey on Value Proposition provided by the Air Transportation startups and companies to estimate their level of expertise and compare it with the pre-training and pre-annotation one. After the participants are explained the startup concept, they learned existing Value proposition major theories. After that they were offered two trainings. Training 1 shows five relatively simple examples of how to determine VP on startups landing pages. Training 2 presents five landing pages where the VP can-

not be easily identified due to its vague expression or the creator of the webpage use non-standard ways to promote their startup, e.g. video or non-trivial words.

The participants have options to identify a landing page as not a startup, can state that the value is hard to identify, and can name the page as not from the air transportation industry. Also, they are asked to click a Like button if they think this startup clearly deliver the VP.

## 4.2 Second Stage – Future Work

**Preliminary results.** In the preliminary results the author observed that the current assumption, based on the suggested algorithm, shows the expected ambiguity of the natural language as well as the task complexity and ambiguity as perception of the VP concept depends on the individual personal characteristics and background. Also, it proves that the necessity of inter-annotator agreement is a measure of how well two (or more) *annotators* can make the same *annotation* decision for a certain category [17]. This moment 56% of the planned startups landing pages are annotated.

**Future Work.** The immediately future work is to identify the features (e.g. linguistic, semantic), so the author can determine specific for startups Value Proposition patterns to develop testable predictions. Creating transportation industry values ontology based on inter-annotated corpus with high ambiguity level is one of the objectives of this research. Building and testing a model is the next phase of the current studies. As the result of this work the author will elaborate a methodology for analysis of values based on a model of automatic identification of values in the text of a startup's landing page in the air transportation industry.

**At the end stage** the author plans to develop a tool in the form of software for assisting industry analysts in their research and advisory services.

Factors that influence the performance of the Value Proposition identifying model are discovered during the annotation process. The author is analyzing participants' comments and commits not only to identify the clearly viewed linguistic features, but also to find not so vivid patterns that we can use to automate the process of VP identification. There are several factors technical and non-technical that the author thinks can influence the performance of annotation process (like what influences the attitude towards and the time of annotation) and that are being investigated.

## Conclusions

In this research the author empirically has studied the process of Value Proposition identification in the air transportation startups. The study has been conducted according to one of the well-recognized path of corpus linguistic research methods: 3A – Annotation, Abstraction, Analysis. The current stage of the given research is a landing text annotation process. An experiment design has been developed: preliminary survey, two-level training, dataset of objects, 100 annotators.

The preliminary findings show the importance of inter-annotator agreement and an ambiguity of natural language and value identification process. The next stage of this research will include identifying the linguistic features and patterns. It will be a basis

of elaborating a methodology of automatic identification of values in the text of a startup's landing page in the air transportation industry. At the end, the author plans to develop a tool in the form of software to assist industry analysts in their research and advisory services.

## References

1. Skytran company webpage (online), [www.skytran.com](http://www.skytran.com), last accessed April 12, 2020
2. McCue ME, McCoy AM. The Scope of Big Data in One Medicine: Unprecedented Opportunities and Challenges. *Front Vet Sci*, e-publishing (2017).
3. Shih W, Sen C. Data-Driven vs. Hypothesis-Driven Research: Making Sense of Big Data. *Academy of Management Proceedings* (2017).
4. von Bertalanffy, L. (1968). *General System Theory: Foundations, Development, Applications*. New York: George Braziller (1968).
5. Burian R.M. Exploratory Experimentation. In: Dubitzky W., Wolkenhauer O., Cho KH., Yokota H. (eds) *Encyclopedia of Systems Biology*. Springer, New York, NY (2013).
6. Ravens C. Methods of statistical analysis. In: *Internal Brand Management in an International Context*. Innovatives Markenmanagement, Springer Gabler, Wiesbaden (2014).
7. Simulation Modeling. In: Alhadj R., Rokne J. (eds) *Encyclopedia of Social Network Analysis and Mining*. Springer, New York, NY (2018).
8. Müller P., Pickard K., Bertsche B. Analysis and Inclusion of Synergies of Common Quality Management Methods for Optimised Quality Assurance. In: Spitzer C., Schmocker U., Dang V.N. *Probabilistic Safety Assessment and Management*. Springer, London (2004).
9. Fu K.S. Syntactic (Linguistic) Pattern Recognition. In: Fu K.S. *Digital Pattern Recognition*. Communication and Cybernetics, vol 10. Springer, Berlin, Heidelberg (1976).
10. Brownlee, Jason. *Discover Feature Engineering, How to Engineer Features and How to Get Good at It*, [machinelearningmastery.com](http://machinelearningmastery.com), (2014)
11. Harvey Golub, Jane Henry, John L. Forbis, Nitin T. Mehta, Michael J. Lanning, Edward G. Michaels, and Kenichi Ohmae. *Delivering value to customers*. Harvard Business Review, President and Fellows of Harvard College (1988).
12. Eric Almquist, John Senior, Nicolas Bloch, "The Elements of Value", *Harvard Business Review*, pp. 46–53 (2016)
13. S. Li, "The Role of Value Proposition and Value Co-Production in New Internet Startups: How New Venture e-Businesses Achieve Competitive Advantage," *PICMET '07 - 2007 Portland International Conference on Management of Engineering & Technology*, Portland, OR, pp. 1126-1132 (2017).
14. Äyväri, Anne; Jyrämä, Annukka: Rethinking. Value proposition tools for living labs. In: *Journal of Service Theory and Practice*, Vol. 27, No. 5, p. 1024-1039 (2017)
15. Guo, Hai, Jun Yang, and Jiaping Han. "The Fit Between Value Proposition Innovation and Technological Innovation in the Digital Environment." *IEEE Transactions on Engineering Management* (2019).
16. Wallis, S. and Nelson G. Knowledge discovery in grammatically analysed corpora. *Data Mining and Knowledge Discovery*, 5: 307-340.(2001)
17. Artstein R. Inter-annotator Agreement. In: Ide N., Pustejovsky J. *Handbook of Linguistic Annotation*. Springer, Dordrecht (2017).