RESEARCH IS A QUEST: reframing a project as a linguistic research program on science communication with the use of data science

Rosana Ferrareto Lourenço Rodrigues
IFSP – São João da Boa Vista

Abstract: *The effects of what science can do depend not just on the knowledge it produces, but on how this knowledge is communicated. For science communicators, one of the challenges is to compress multi-dimensional information into language that is accessible and meaningful to their target audiences. In order to explore ways to broaden the foundations of studies on Science Communication, we state that there is a methodological need to test hypothesis against data. By this approach, we mean to reframe a project in Cognitive Linguistics for teaching scientific writing through introspection as a program whose investigation is based on a massive systematic corpus of ecologically valid multimodal data, along with tools and practices to analyze this data. This Data Science approach is found in The Distributed Little Red Hen Lab, where we have developed the Basic Text Pipeline to investigate narrativity and figurativity in Science Communication from a corpus of 1,000 journal abstracts. The outcomes are to present theoretical contribution from Cognitive Linguistics to Science Communication, to model it from frame narratives and blending constructions and to develop tools for research purposes on this discipline as well as for teaching and learning scientific writing.*

Keywords: *Science Communication. Cognitive Linguistics. Narrativity. Figurativity. Data Science. Scientific Writing.*

"Communication is not something you add on to science; it is of the essence of science"

(Alan Alda).

Science challenges conventional thinking. By exploring knowledge, scientists cross boundaries, find new patterns, unlock the unknown. Science communication can be powerful enough to change the natural and social world. It can enhance quality of life. The effects of what science can do depend not just on the knowledge it produces, but on how this knowledge is communicated.

For science communicators, one of the challenges is to compress multi-dimensional information into language that is accessible and meaningful to their target audiences. Effective Science Communication involves frames to highlight scientific findings. Scientific stories should be conceptually straightforward as to explain a problem and describe a solution. The structure PROBLEM → SOLUTION can be unfolded as a Quest, framed as an Attempt to achieve a Goal.

In the story of Scientific Research framed as a Quest, one can picture the scientist (the Agent) frequently raising questions, coming up with a plan, facing obstacles that hold back the routine of an experiment (the Attempts) before finally coming to a resolution built on that knowledge to create something new (the Goal). This is the higher (cognitive) narrative that scientists should be able to communicate, not only by expressing this path logically but also by unfolding it in language creatively.

With this context in mind, our aim is to model Science Communication from frame narratives and blending constructions in order to present a Cognitive Linguistics approach for teaching and learning scientific writing. We argue that frames are a core component of human cognition, crucial for communicating meaning via shared image schemas and conceptual integration.

In order to explore ways to broaden the foundations of studies on Science Communication, we state that there is a methodological need to test hypothesis against data. By this approach, we mean to reframe a project in Cognitive Linguistics for teaching scientific writing through introspection as a program whose investigation is based on a massive systematic corpus of ecologically valid multimodal data, along with tools and practices to analyze this data. This Data Science approach is found in The Distributed Little Red Hen Lab, where we have developed the Basic Text Pipeline to investigate narrativity and figurativity in Science Communication from a corpus of 1,000 journal abstracts.

This is our point of departure: using Frame Semantics to model the scientific cognitive narrative through Automatic Frame Annotation in order to look at conceptual structure patterns; and using a Corpus WorkBench and Query Processor to track productive linguistic blending constructions inside the scientific text genres. Not only are we interested in understanding how scientists frame their research process logically when they write papers but also we want to look at how they develop, describe and communicate scientific concepts creatively.

Our destination is to contribute to Cognitive Linguistics Research and unfold a new story for Science Communication by achieving valid and reliable results with pedagogical and technological applications.

1. Setting the scene: Science Communication as a discipline

Science is valued by society because the application of scientific knowledge helps to fulfill many basic human needs and improve living standards. Directions and benefits out of scientific findings are expected to be shared. That is why communicating science should be the sole purpose of doing science. Science's ability to have any effects depends not just on the content of research activity but also on how effectively this content – which is often-complex information – is communicated (DRUCKMAN; LUPIA, 2017).

There has been a long tradition of popularizing science communication as for disseminating scientific discoveries to the general public. In this context, Science Communication taken as a discipline in Academia is supposed to empower scientists to communicate complex topics in clear, vivid, and engaging ways; leading to improved understanding by the public, but also by the media, and others outside of their own discipline (ALDA, 2019).

By setting the scene of Science Communication as a discipline, we aim to trace back to its history and shed some light on its recent changing landscape. We want to look at its outreach not only to the general public, but most importantly to the first ones to filter information and make a decision for publication – the journals´ reviewers. Also, we are not only concerned about the general public´s interpretation of published papers. In multidisciplinary environments and in times when interdisciplinary research is demanded, the specialized audience of scientists have to be able to understand complex scientific concepts from various fields.

It is important to point out that scientific writers´ needs and paths are diverse. Students need to write a thesis, a faculty member needs to compose a grant proposal, or a public information officer needs to craft a press release. There is now an increased focus on those writers working in corporate settings, government, and nonprofit organizations. Scientists need writing for translating scientific material, due to the globalization of research. They need to address controversial issues such as climate change or emerging viruses. They need to be able to move seamlessly among platforms and styles. Scientists and researchers need to be able to expertly connect with their audiences, no matter the medium (MONTGOMERY, 2017).

This rapid switch to expanding area for creating meaningful engagement between scientists and this diverse target audience is evident these days. Ngumbi (2018) states that the appetite for Science Communication is on the rise. In a short span of time, several workshops for students, postdocs, early career and seasoned scientists have taken place at institutions around the world and scientists want to "step out of the ivory tower to share what they know with the public" (BARON, 2010).

Although universities, research institutions and professional societies are stepping up their efforts to train and equip scientists with resources they need to effectively disseminate scientific

discoveries with the public, Science Communication remains a relatively new field to many science majors (NGUMBI, 2018).

This has opened up space for various resources and initiatives such as guides to the design and delivery of effective engagement practices to help novice scientists and even to a certified course in Science Communication[1] (BOWATER; YEOMAN, 2012). There is also room for new trends involving Science Communication as a discipline such as the one of storytelling[2]: scientists are considered to be artists and storytellers, and "story is about having something really cool you want to share with your audience, and then putting it in a context where that is valuable and unknown. And then people go with you to find it out" (KERBY in ALDA, 2019). Besides, there is an increasing number of events for science communicators[3] and writing centers at universities around the world[4] so that scientists can learn to do Science Communication and publish papers rather than only talking about it.

Although there are scientific studies on Science Communication[5], mostly in Social Sciences in disciplines such as Politics and Economics, what still remains almost untouched is the need to tailor approaches based on scientific investigations on Science Communication in different areas of study not only to communicate with the general public, but specially to understand Science Communication in Academia in interactions among different fields of knowledge.

What is missing in this scenario is an investigation on how language affects the reviewer´s (a scientist who is not always from the same discipline/field) evaluation and thus influence publication. This refers back to how important it is for science communication to be effective. Effective science communication often involves frames that highlight particular aspects of a scientific finding or issue (DRUCKMAN; LUPIA, 2017).

We argue that Cognitive Linguistics can provide Science Communication with constructs through which we can look at these particular aspects from linguistic and cognitive realms.

2. The territory of Cognitive Linguistics: blending, frame, narrative, image schemas and conceptual metaphor

"Science is the mother of technology" is usually stated in institutions where science is done with the use of technology. This linguistic construction – The X is the Y of Z – in which X = science,

1 This Certificate in Science Communication is offered by the Beckman Institute of the University of Illinois, jointly by the 21st Century Scientists Working Group and the Center for Innovation in Teaching and Learning. It is designed for graduate students, but also open to postdocs (<https://21centurysci.beckman.illinois.edu/science-communication-certificate/>).

2 There is a model of story design based on shows and writing plays and on teaching about visual communication and effective data visualization (<https://news.wisc.edu/new-course-brings-storytelling-techniques-to-science/).

3 See: The greatest thing ever to happen to science communication (<http://www.deepseanews.com/2017/11/the-greatest-thing-ever-to-happen-to-science-communication/>).

4 See: Write Well or Perish (<http://revistapesquisa.fapesp.br/en/2011/04/01/write-well-or-perish/>).

5 See: <https://journals.sagepub.com/home/scx> and <https://jcom.sissa.it/about>.

Y = mother and Z = technology in an analogy first noted by Aristotle and studied in details by Turner (1997). Science is to mother as technology is to daughter (the element W = daughter is inferred from the value identified in mother). Other examples would be: "Vade Mecum[6] is the Wikipedia of Law" and "Tanenbaum[7] is the Bible of computer architecture". This kind of linguistic construction is didactically persuasive as it anchors new information (technology, Vade Mecum and Tennenbaun) to previous knowledge (science, Wikipedia and the Bible). In just a few words in a single sentence, elements of different scenes/domains are "packed together" to unveil the new through the logic. This is one of the bullet points on the map for the territory of Cognitive Linguistics: scenes "packed together" are frames (FILLMORE, 1982) and the conceptual integration that glues the packs is blending or frame blending (TURNER, 2008A; FAUCONNIER; TURNER, 2002).

Blending refers to meaning construction that involves integration of structure that gives rise to more than the sum of its parts (FAUCONNIER; TURNER, 2002). In Science Communication, it is a mechanism for managing complex ideas by integrating two types of knowledge and create a third new unprecedented one. In "Tanenbaum is the Bible of computer architecture", we integrate two frames: "computer architecture" (of a computational domain) and "the Bible" (of a religious domain). The new meaning is Tanenbaum´s publication seen as the holy book by computer scientists.

Frame is a knowledge structure represented at the conceptual level and held in long-term memory and which relates elements and entities associated with a cultural scene, situation or event from human experience (FILLMORE, 1982). In Science Communication, the same concept can be described in different ways depending on the writer´s purpose and the angle from which the concept is framed. For example, in a paper of Linguistics, one can find a given result discussion about Semantics described as "This is a well-crafted sentence meaning. You should form your words carefully. Ideas should be articulated clearly". Objects created or altered for some purpose is the concept being communicated here under the frame Artifact[8]. The frame Crafting makes use of Artifact (ideas, words, sentences) and is a subcase of Purposeful action (communicating/ articulating meaning clearly). It has as a source frame the conceptual metaphor FORMING WORDS IS SHAPING, which is an entailment of WORDS ARE CONTAINERS (MetaNet, 2018)[9].

The communication process happens when scientists describe concepts in the text to evoke such an interpretation, which is related to the linguistic realm of Science Communication.

The research process is also invoked in the reader´s mind when concepts are framed by scientists. This is in the cognitive realm of Science Communication. For example, a particular Activity, performed conventionally or habitually by more than one Practitioner within a Culture, is described in terms of its method. These stages and elements are represented under the frame <Craft> (FrameNet, 2019). If we frame the concept Science as Craft we can picture literature reviews and scientific experiments (Activity) being conducted from test to results (habit, method, convention) by

6 <https://www.merriam-webster.com/dictionary/vade%20mecum>.
7 <https://books.google.com/books/about/Structured_Computer_Organization.html?id=EREwDwAAQBAJ>
8 We have searched for frames in FrameNet. Information about what FrameNet is can be found on: < https://framenet.icsi.berkeley.edu/fndrupal/WhatIsFrameNet>.
9 Information about what MetaNet is can be found on: < https://metanet.icsi.berkeley.edu/metanet/>.

researchers/scientists (Practitioners) in labs or in the field within Academia (Culture) described in a published paper (method, convention).

Frames are powerful tools for developing narrative patterns. A narrative is a recordable sequence of events associated with a given topic as a means of transferring information, experience, attitude or point of view. By adding characters, places, time, actions and events, narration and focalization, we create a story, a cognitive artifact used by humans for thinking, feeling, perceiving and understanding (HERMAN, 2003). In Science Communication, narrative functions as a cognitive and communicative mode for supporting a variety of problem-solving activities including the ones underlying the research process.

One of the constructs that helps to construct a narrative structure in scientific texts is image schema. Image schemas are abstract conceptual representation that arises from our everyday interaction with and observation of the world around us. They derive from sensory and perceptual (embodied) experience (JOHNSON, 1987). For example, in Science framed as Craft, words are CONTAINER-like Artifacts. Also, the physical movement along a path uses the image-schematic structure SOURCE-PATH-GOAL. In Research framed as a Quest, when we think about and visualize the story of scientists coming up to a solution by unfolding a problem through the use of methods, we are thinking image-schematically. That means that we are using the SOURCE-PATH-GOAL schema to structure the concept of scientific research and to think of the method (PATH) as the means between the problem (SOURCE) and the solution (GOAL) . This image-schematic structure is crucial for narrating the story of research. It consists of movement through time along the path of science. In doing science for communication this helps scientists to frame (the format of) a paper logically. Frames are a core component of human communication, crucial for communicating meaning via shared schemas. Indeed, the choice of a particular frame can be the key to conveying vital scientific information effectively (DRUCKMAN; LUPIA, 2017).

Image schemas are a conceptual representation of spatial relations that are metaphorically used in non-spatial domains to scaffold abstract ideas, such as the one of WORDS ARE CONTAINERS evoked in the linguistic material. In addition, the structure of Research as a Quest is used from the physical domain of the frame Journey (whose elements are a path, a traveler, a vehicle, a map, a territory) to scaffold abstract concepts, such as the process of changing over time entailed by the scientific concept of science as method, as we have illustrated in the example of scientists facing up obstacles before coming up to results out of questions raised from a problem. This is invoked in the researcher´s mind.

Time mapped in terms of space consists of a conceptual metaphor: a conceptual projection involving mappings between distinct conceptual domains – a source domain (concrete) and a target domain (abstract) (LAKOFF; JOHNSON, 1980). In Science Communication, these are mappings between knowledge from the domain that is the object of comprehension (still abstract), and knowledge from another distinct domain (already concrete) that one uses to construe this comprehension. In the example of “having well-crafted sentences for clear meaning”, artifacts

such as words as containers (concrete) are used for describing clear communication (abstract). The conceptual metaphor CREATING LINGUISTIC EXPRESSIONS IS SHAPE MANIPULATION has the Source frame Crafting and the Target Frame Forming words (MetaNet, 2018), mapped and blended so that the concept of language as Artifact and communication as Craft are described to communicate the idea of "how to make clear meaning when writing or speaking".

Frame blending prompts for compression and blending of two distinctive narratives, and can be highly metaphorical (TURNER, 2003, 2008a, 2008b). We usually trace and locate two distinct frames in the same sentence or excerpt in scientific texts. From the integration of seemingly distinct and new abstract concepts, one obtains an accessible understanding of the new information based on prior concrete knowledge, for example, understanding "Vade Mecum" (in the frame of Law) in terms of "Wikipedia" (in the frame of Internet). Another example is the WAR and DISEASE frames in sentences about cancer in Medicine articles, such as "The patient has lost the battle against cancer". Another example would be "Editing the human genome is playing with the software of life," in which the concept of evolution through genetic engineering is mapped in terms of software engineering processes. In Science Communication, framing can produce better learning outcomes (DRUCKMAN; LUPIA, 2017).

The relevance of frame blending for Science Communication lies on how concepts of distinct fields are integrated for interdisciplinary research, such as Biotechnology. This facilitates the understanding of abstract scientific concepts by scientists from other fields and also by the public since frame blending, i.e. conceptual integration via frames, makes communication more effective. When we explain a relatively complex idea in a new way, we reduce the complexity of the idea by blending and compressing – over time, space, causation, and agency – vital conceptual relations between different mental spaces making outer-space relations becoming inner-space relations (FAUCONNIER; TURNER, 2002; TURNER, 2006; 2017).

In a nutshell, in (Science) Communication, framing, metaphor, image schemas, narrative cognition and blending "characterize human higher-order cognition, with its species-wide capacities for exceptional creativity and innovation. […] They labor together and result in a feedback loop – an autocatalytic vortex, a self-reinforcing cyclone of cognitive innovation" (Turner, 2008b, p. 2-3).

The challenge facing science communicators is that the scientific process is itself a complex phenomenon:

> A researcher chooses where, when, and how to gather evidence. A researcher also chooses what metrics to use to characterize observations. A researcher chooses how to analyze the observations. In many cases, researchers choose a particular statistical model. Attempts to understand the full meaning of a scientific finding can depend on knowledge of how the finding was produced (DRUCKMAN; LUPIA, 2017, p. 2).

Also, science communicators must decide what aspects of the topic and research design to describe first and which aspects to convey later:

> Science communicators who make these choices are involved in acts of compression. They are seeking a means of converting high-dimensional research phenomena and multifaceted research processes into language that is accessible and meaningful to their target audiences (DRUCKMAN; LUPIA, 2017, p. 2).

Compression (analogy towards identity) as well as decompression (disanology towards change) (FAUCONNIER; TURNER, 2002) between the research process and the communication process are cognitive mechanisms not only for organizing the research story in a paper logically, but also for developing scientific concepts manifested in linguistic constructions[10] creatively. The underlying cognitive narrative functions as a communicative mode for supporting a variety of problem-solving activities in the research process, which are materialistically anchored in language in the communication process.

Those theoretical reflections have guided us in the territory of Cognitive Linguistics towards the investigation of Science Communication. We aim to go both ways: from language to cognition and from cognition to language so that we can guide scientists to focus on relevant processes and concepts from their research in order to make effective communicative choices when writing papers. This path is visualized in the conceptual model we have designed in order to account for narrativity and figurativity in scientific texts.

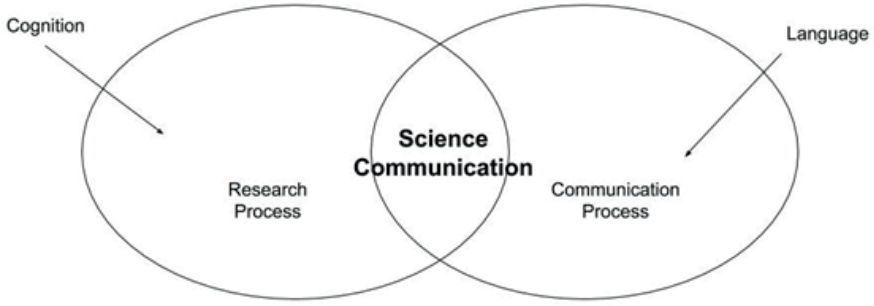
3. The roadmap: modeling narrativity and figurativity in Science Communication for teaching scientific writing

Science Communication entails both Research as a Quest and Science as Craft: it is a blend between how researchers frame their practice and how they express it when writing papers. It lies in between the Research Process and the Communication Process (Figure 1).

The research narrative anchors the linguistic development and the linguistic development is anchored to the research narrative. One way to access the cognitive process of research, which is invoked in the reader's mind and primes text interpretation, is by analysing the linguistic material. How researchers narrate knowledge is a reflection of how they understand that knowledge and thus how they develop, describe and communicate scientific concepts, which are evoked by the linguistic material: from the choice of given frames for each stage of the research (narrativity) to the use of given linguistic constructions (figurativity).

10 Form-meaning pairs are often in Linguistics called "constructions" (See Fillmore et al. 1988).

Figure 1 – Science Communication: cognition/research & language/communication



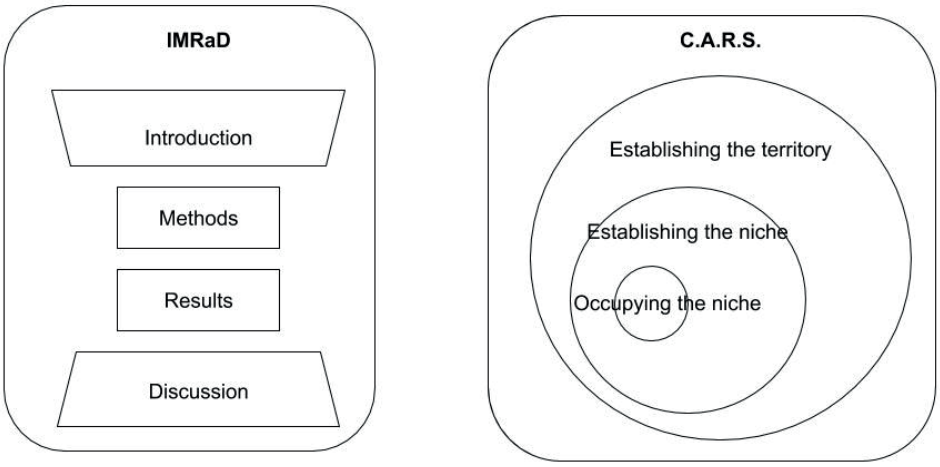
Source: Own creation

The underlying cognition in the use of language for communication make up the narrative and the figurative structures in Science Communication. The scientific process at the heart of this intersection between Research and Communication requires knowledge of context (field), content (topic) and language (rhetoric and text genre) (DRUCKMAN; LUPIA, 2017; MCCUTCHEN; TESKE; BANKSTON, 2008; SWALES, 1990; 2004; SWALES; FEAK, 2009); and skills on critical thinking, decision making and problem solving for discussing and questioning ideas fluently, based on evidences, creative thinking, reasoning and logical inferencing (MEMIŞ; ÖZ, 2014). "The combinations of context and content can help science communicators more effectively use frames to convey critical information to important audiences" (DRUCKMAN; LUPIA, 2017, p. 11).

In the writing (cognitive) process, knowledge consists of two sources: deep knowledge of the topic and deep knowledge of the text genre. In Science Communication, while the first is achieved through literature review, the second is primarily achieved by walking through the structure <Introduction – Methodology – Analysis & Results – Discussion & Conclusion>. It is the knowledge of the scientific text genre that structures content and assists the ongoing research process (MCCUTCHEN; TESKE; BANKSTON, 2008).

The structure accounts for the wineglass model named IMRaD, which stands for Introduction, Methods, Results and Discussions, which are the sections in a manuscript format (CARGILL; O'CONNOR, 2008). The rhetorical genre "scientific articles" is found in a framework proposed by Swales (1990; 2004; SWALES; FEAK, 2009) for writing introductions and abstracts from moves, named C.A.R.S., which stands for Create a Research Space from establishing a territory and establishing a niche to occupying the niche. According to this model, in establishing a territory, scientists should make a claim, review previous work and position themselves. In establishing a niche, they raise a problem, indicate a gap or continue developing an existing tradition. In occupying the niche, they show the importance of their work by presenting a purpose and providing an outline of their research. The models for scientific writing IMRaD and C.A.R.S. (Figure 2) respectively bring about the Knowledge of the genre (language) and knowledge of the topic and the field (content and context).

Figure 2 – Models for scientific writing IMRaD and C.A.R.S.



Source: Own creation

Starting from the perspective of these models but in order to advance and create a roadmap for scientists to write papers and communicate their scientific research design and concepts more effectively, our proposal is to model the narrativity and figurativity in Science Communication for teaching scientific writing. We adopt the approach of writing as a (learning) process as opposed to a product-based approach.

Scientific writing is traditionally taken as a verbal production skill whose instruction is primarily focused on particular types of structure and style, capturing the compositional aspects of scientific texts required for content production and communication, putting aside the cognitive aspects involved in the writing process (MEMIŞ; ÖZ, 2014).

In the process-based approach, writing is taught as a learning tool: a vehicle to access and express a thought; a device to evaluate and interpret a piece of thinking; a means to build connections in a text; a planning mechanism for thinking ahead backwards (MEMIŞ; ÖZ, 2014). Some of the bullet points in this map will have scientists as writers get to know how to generate ideas and transform them into text; how to think logically and creatively; how to promote conceptual understanding of Science, which is a complex cognitive activity, as a frame narrative through blending linguistic constructions as material anchors.

The Cognitive Linguistics constructs of blending, frames, metaphor, image schemas and narrative have been added to these models of composition and style within this approach of writing as a (cognitive) process. These hypothesis have been triggered by the fact that although most scientists are able to write papers in the expected format, they can´t still get them published.

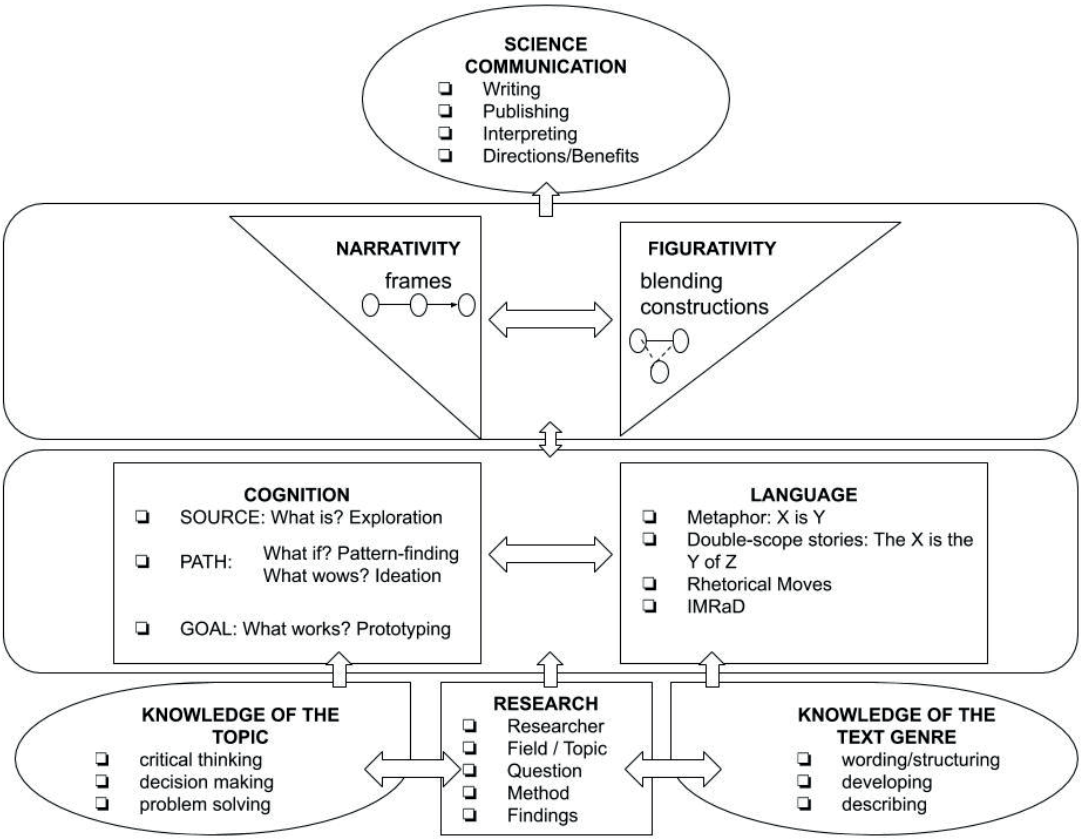
Our first attempts were around modeling the underlying narrative of a paper in a 3-act (moves in C.A.R.S. by SWALES, 1990) image schematic structure SOURCE-PATH-GOAL by thinking of the frames Research and Science to visualize a script where Science is Craft and Research is a Quest as well as by coming up with a Frame List from FrameNet (2019) according to IMRaD (CARGILL; O'CONNOR, 2008), expecting to find linguistic constructions which would evoke the following frames in each of the paper sections: (i) Introduction: Questioning, Causation, Purpose, Perception_ experience, Perception_active, Complianc, Becoming_aware, Development; (ii) Method: Trying_ out, Means, Examination, Experimentation, Intentionally_act, Intentionally_create, Organization, Pattern; and (iii) Discussion and Results: Being_in_effect, Evidence, Certainty, Trust, Coming_to_ believe, Resolve_problem, Cause_benefit_or_detriment, Response, Using, Direction.

After refining these first attempts, we have come up with a model for narrativity & figurativity in Science Communication (Figure 3), using the Design Thinking framework as a leverage point. Design Thinking is a systematic approach to problem solving and decision making, taught to managers as a process and everyday tool for profitable growth to drive innovation in their companies (LIEDTKA; OGILVIE, 2011). Tool 1 of Design Thinking is Visualization:

> Visualization consciously inserts visual imagery into our work processes and focuses on bringing an idea to life, simplifying team collaboration and (eventually) creating stories that go to the heart of how designers cultivate empathy in every phase of their work and use it to generate excitement for new ideas (LIEDTKA; OGILVIE, 2011, p. 492-494).

Science is about using logic and creativity towards problem solving and decision making and that is why we have chosen Design Thinking. It is a creative and learning process, usually used to tackle complex, sometimes unknown, problems by focusing on users (scientists) and their needs (unfolding problems into solutions). We have chosen Visualization as a tool for building a model on Science Communication because one of our aims is pedagogy: we want to use it to teach scientific writing and have scientists achieve successful publication.

Figure 3 – Science Communication Model



Source: Own creation

By modeling narrativity and figurativity from the research and the communication processes in Science Communication, we start by framing Research: there is a researcher within a field of knowledge and an investigation on a given topic which starts from a question. There is a choice for a method which will lead to findings. In the process, two kinds of knowledge are needed: knowledge of the topic and knowledge of the text genre "scientific paper". Cognition and language are at play for critical thinking, decision making and problem solving and writing starts as a learning process of wording/structuring, developing and describing scientific stages and concepts for communication and publication.

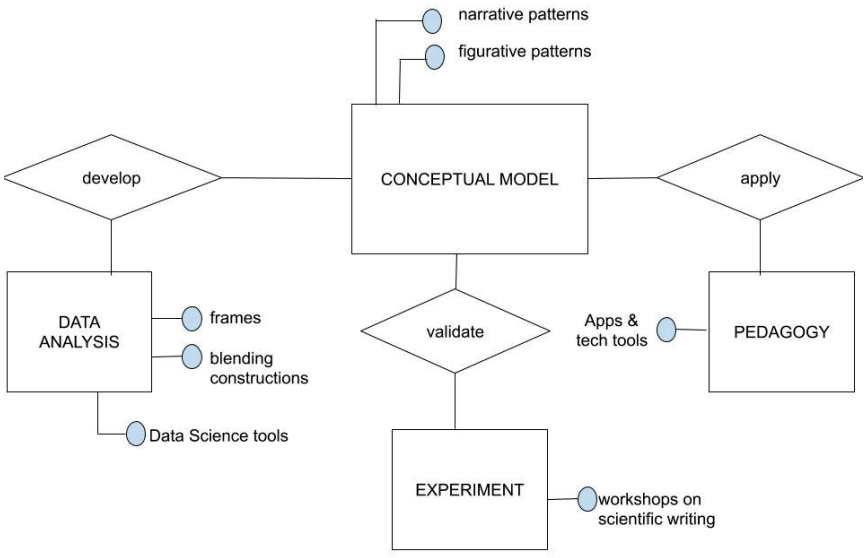
Thinking of the research stages logically and creatively consists of a move from SOURCE to GOAL. We model the PATH with the following Design Thinking tools: it starts with an Exploration stage, in which scientists tackle the question *What is*, a data-based exploratory question for current reality. After identifying and defining the problem, scientists initiate Pattern-finding and tackle the question *What if*. They brainstorm possible answers to the question, by imagining ways to solve the problem. Next there is Ideation, tackled by the question *What wows*. There is a move from a hypothesis-generating mode to a testing mode, by resorting to experiments and tests of the assumptions underlying each hypothesis. Scientists make some choices. Finally, Prototyping tackles the question *What works*. In this launch and learning stage, scientists activate their creating skills to

offer a prototype of their idea as a solution to the problem (LIEDTKA; OGILVIE, 2011).

To understand human communication, we must develop a new facility for understanding the grammar of meaning construction and cognitive analysis of communicative practices (STEEN; TURNER, 2013). Thus besides organizing the cognitive narrative of their research into steps of a process, scientists have to register it in a text into sections of a paper. They structure it in the expected conventional format (IMRaD) and conform it to the rhetorical genre composition and style (C.A.R.S.). In order to develop and describe concepts with creativity, they can word them by using blending constructions such as metaphor and frame blending, which accounts for figurativity in Cognitive Linguistics constructs. Narrativity from frames and figurativity from blending constructions can be analyzed in texts using technological tools developed out of the model, which will be presented in Section 4 of this paper. The final step for Science Communication is to have the paper written and published so that readers can benefit from directions through interpretation.

We plan to go from Model to Methodology (Figure 4) by: (i) applying this model for teaching scientific writing with the development and use of technological tools; (ii) testing and validating it in experiments in workshops for teaching scientific writing; and (iii) developing it through data analysis with the use of Data Science tools.

Figure 4 – From Model to Methodology



Source: Own creation

The methodological need to test hypothesis against a massive systematic corpus of ecologically valid multimodal data, along with tools and practices to analyze this data is found in the The Distributed Little Red Hen Lab[11].

11 Red Hen (for short) < http://www.redhenlab.org/>.

Moving on from theory and model and from model to tools – the investigation of frames in the Scientific Domain with the use of semantic parsers such as Open Sesame and Semafor based on FrameNet; and the investigation of blending constructions with the use of Parts of Speech (POS) tagger in Standford Core NLP, both in a Corpus WorkBench and Query Processor (CQPWeb) – will have us see the narrative and figurative patterns in Science Communication.

4. Moving on: Red Hen Big Data Science Tools for Linguistics Research

"Maybe stories are just data with a soul" (Brené Brown).

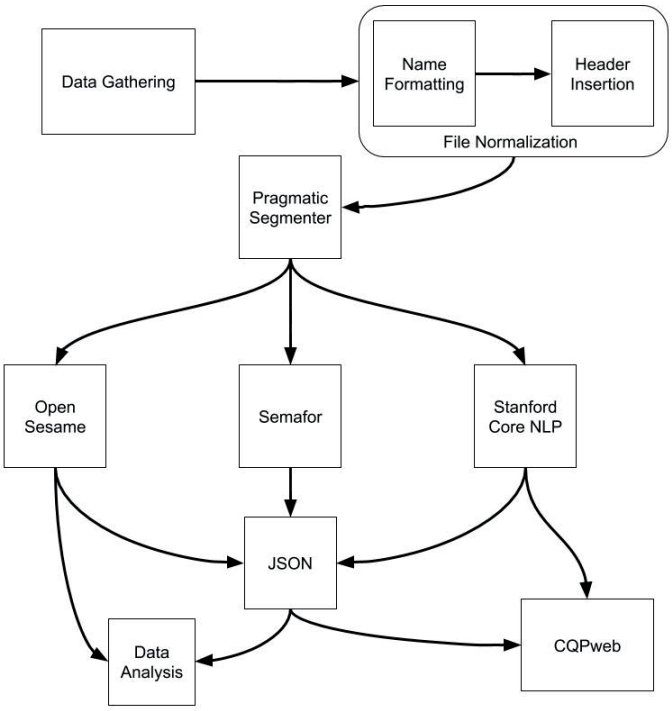
Data Science[12] tools and practices open up a new horizon to Linguistics Research. In the field of Linguistics, there are traditional methodologies which are already standardized such as introspection with theoretical analysis (TALMY, 2006). Red Hen can be used to search for productive patterns in the use of language, which can advance the analysts´ attention to particular aspects of language as manifested in the language user´s cognition and thus help validate hypothesis from data.

Red Hen is a corpus of closed captions, transcripts, on-screen text and now a prose corpus of scientific abstracts. It is additionally ready to export all her hits to a comma-separated-value file, which can then be fed to a software package like R[13], and subjected to statistical analysis.

For this study on Science Communication, a corpus[14] of 1,000 abstracts from high impact journals in ten different disciplines in the Web of Science[15] was compiled and it is in the Basic Text Pipeline[16] (Figure 5), in the website barnyard of Red Hen.

12 <https://en.wikipedia.org/wiki/Data_science>.
13 < www.r-project.org>.
14 Information about criteria and composition of the corpus can be found on < https://drive.google.com/file/d/1EGO-78sN1KA15MbG-FvT--riASxFBQDWZ/view?usp=sharing>.
15 <https://www.webofknowledge.com/>.
16 <http://www.redhenlab.org/home/the-cognitive-core-research-topics-in-red-hen/the-barnyard/basic-text-pipeline>.

Figure 5 – The Basic Text Pipeline



Source: Own creation

The diagram in Figure 5 shows how Brazilian Red Hens[17] are working on tools so that we can have the abstract corpus annotated by frames by Open Sesame[18] and Semafor[19] and by POS from Standford Core NLP[20] and thus searchable for narrative frames and blending constructions in CQPweb[21].

After data gathering (the abstract corpus compilation), file normalization was done so that we could have the texts ready for sentence splitting in the Pragmatic Segmenter, since the semantic parsers OpenSesame and Semafor annotate frames sentence by sentence from target (lexical units which evoke frames), frames and then arguments (frame elements). To achieve automatic frame annotation, we can use both Semafor 3.0, which works with FrameNet 1.5, and OpenSesame, which also handles FrameNet 1.7.[22]

In addition to frame annotation tools (Semafor and OpenSesame), we are using a query processor in a Corpus WorkBench (CQPWeb) to search for constructions such as X is Y for metaphors and X is Y of Z for frame blending. Corpora are annotated by POS, that is, tagged morphologically, using the Stanford Core NLP so that queries can be made out of the grammatical patterns such as:

17 Brazilian Red Hens are IFSP (Federal Institute of Education, Science and Technology of Sao Paulo) undergraduate students in Computer Science.
18 < https://github.com/swabhs/open-sesame>
19 < http://www.cs.cmu.edu/~ark/SEMAFOR/>
20 <https://stanfordnlp.github.io/CoreNLP/>
21 <https://cqpweb.lancs.ac.uk/>
22 See Butterfly Effects in Frame Semantic Parsing: impact of data processing on model ranking (2018) <aclweb.org/anthology/C18-1267>.

_NN is _NN and _NN is _NN of_NN[23]. The POS annotation will allow us to then make searches and look at distributions based on the metadata present defined as categories that we are interested in for analysis, such as: text id/file; abstract title; publication_venue (or journal_name); publication_year; abstract content (or abstract_body or abstract_text); discipline and field. These are the variables of the data analysis, whose research questions are to see what narrative and figurative patterns are in the annotated corpus of scientific abstracts and to investigate in what extent they can cause a paper to get published (or not) in high impact journals and why.

The Brazilian Red Hen team is working on the development of an integrated application in a JSON[24] structure so that we can benefit, for our analysis, from both Open Sesame/Semafor and Standford Core NLP and have the corpus annotated by both frames and POS within CQPweb in such a way that we can search for narrative (from frames) and figurative patterns (from linguistic constructions).

Once data analysis is done with the use of these tools, and by using R statistical package with frequency lists and graphs, we may be able to visualize what the prototypical frames and constructions are in this corpus of scientific abstracts published in high impact journals and thus propose a taxonomy which can be used for teaching scientific writing for successful publishing.

Future applications are on the way after this data analysis out of the Basic Text pipeline is carried out for the following purposes (see Figure 4): for developing the model we propose (see Figure 3), for validating it in an experiment and for developing other tools for research and pedagogical purposes on working with Science Communication.

5. Towards a new Quest in Science Communication: theoretical, pedagogical and technological paths

Science Communication lies between the research process and the communication process, and this study proposes to show how language can affect/be affected by cognition and influence/be influenced by communication.

What is the meaning of these preliminary results in this study of frame narratives and blending constructions in Science Communication? Theoretically, there are implications for metaphor research in the extent that we can understand what is conceptually salient in the minds of scientists when they develop and communicate concepts. There may be contribution to FrameNet by adding new specific frames to the scientific domain. Pedagogically and technologically, the model resulting from this study can help didactic practices and the development of technological tools for the teaching of scientific writing and for research on Science Communication.

What is next? Frame and linguistic construction statistics out of the corpus annotated by

23 _NN is for noun in the tagset used - the Penn Treebank <https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html>.

24 <https://www.json.org/>

OpenSesame/Semafor/Standford Core NLP and within CQPweb, and visualized through R statistical package, will provide us with a picture of the narrative and figurative patterns in published papers. This can validate our model on Science Communication, which can be used to teach scientific writing for successful publishing.

This has turned into a new Quest: a project has become a sustainable research program whose new dimensions can bring theoretical contributions to Cognitive Linguistics and potential pedagogical & technological applications in Science Communication.

This is not the end of the story on how Cognitive Linguistics can make Science Communication more effective. This is the beginning of a whole new story in an international transdisciplinary collaboration with Red Hen for future development of new techniques in the barnyard, such as tagging for story structure and conceptual blends via image schemas.

Acknowledgements: This study was partially supported by the Federal Institute of Education, Science and Technology of Sao Paulo (IFSP) under the Research Grant n. 2018/2.348; and by Sao Paulo Research Foundation (FAPESP) under the Research Grant n. 2017/15988-2, for a Postdoctoral Research (2018-2019) at Case Western Reserve University, Cleveland, OH – USA, in the Department of Cognitive Science, under the supervision of Dr. Mark Turner.

References

ALDA, A. *Alda Center for Communicating Science*. New York: Stony Brook University, 2014-2019. Disponível em: < https://www.aldacenter.org/get-started/about-us>. Acesso em: 2 Jul. 2019.

BARON, N. *Escape from the Ivory Tower:* a guide to making your science matter 2nd edition. Washington: Island Press, 2010.

BOWATER, L.; YEOMAN, K. *Science Communication:* a practical guide for scientists. New Jersey: Wiley-Blackwell, 2012.

CARGILL, M.; O'CONNOR, P. *Writing Scientific Research Articles:* Strategy and Steps. Wiley-Blackwell, 2008.

DRUCKMAN, J. N.; LUPIA, A. Using Frames to Make Scientific Communication More Effective. *The Oxford Handbook of the Science of Science Communication.* Edited by Kathleen Hall Jamieson, Dan M. Kahan, and Dietram A. Scheufele, 2017.

FAUCONNIER, G.; TURNER, M. *The Way We Think:* Conceptual Blending and the Mind's Hidden Complexities. New York: Basic, 2002.

FILLMORE, C; KAY, P; O'CONNOR, M. C. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language* 64(3), 1988, p. 501–538.

FILLMORE. C. Frame Semantics. In: *Linguistics in the morning calm.* Seoul: Hanshin, 1982, p.111-138.

FRAMENET, Berkeley. 2019. Disponível em: <https://framenet2.icsi.berkeley.edu>. Acesso em: 2 Jul. 2019.

HERMAN, D. *Narrative Theory and the Cognitive Science* (ed.). Standford, California: The University of Chicago Press Books, 2003.

JOHNSON, M. *The Body in the Mind:* The Bodily Basis of Meaning, Imagination, and Reason. Chicago: University of Chicago Press, 1987.

LAKOFF, G.; JOHNSON, M. *Metaphors we live by.* Chicago: The University of Chicago Press, 1980.

LIEDTKA, J.; OGILVIE, T. *Designing for growth:* a Design Thinking tool kit for managers. New York: Columbia University Press, 2011.

MCCUTCHEN, D.; TESKE, P.; BANKSTON, C. Writing and Cognition: Implications of the Cognitive Architecture for Learning to Write and Writing to Learn. In: Charles Bazerman (ed.) *Handbook of Research on Writing: History, Society, School, Individual, Text.* New York: Taylor & Francis Group, 2008. p. 574-598.

MEMIS, E. K.; ÖZ, M. The Impact of Inquiry Process on the Cognitive Process Dimensions of Nontraditional Writing. *Mediterranean Journal of Social Sciences*, [S.l.], v. 5, n. 20, 2014. p. 1158.

METANET Metaphor Wiki, Berkeley. 2018. Disponível em: < https://metaphor.icsi.berkeley.edu/pub/en/index.php/MetaNet_Metaphor_Wiki>. Acesso em: 2 Jul. 2019.

MONTGOMERY, S. L. *The Chicago Guide to Communicating Science*. 2nd edition. Chicago: University of Chicago Press, 2017.

NGUMBI, E. *Scientists are learning the power of outreach:* progress in conveying science to the public needs to continue. Scientific America, 2018. Disponível em: <https://blogs.scientificamerican.com/observations/scientists-are-learning-the-power-of-outreach/?redirect=1>. Acesso em: 2 Jul. 2019.

STEEN, F; TURNER, M. Multimodal Construction Grammar. *Language and the Creative Mind.* Borkent, Michael, Barbara Dancygier, and Jennifer Hinnell (eds.).

CSLI Publications, 2013.

SWALES, J. *Genre analysis:* English in academic and research settings. Cambridge University Press, 1990.

SWALES, J. *Research genres:* Exploration and applications. Cambridge: Cambridge University Press, 2004.