TRADITION AND TECHNOLOGY:

A DESIGN-BASED PROTOTYPE OF AN ONLINE GINAN SEMANTIZATION TOOL

A Dissertation Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy
In the Department of Curriculum Studies
University of Saskatchewan
Saskatoon

By

KARIM KAMIL THARANI

PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

DISCLAIMER

Reference in this dissertation to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this dissertation in whole or part should be addressed to:

Head of the Department of Curriculum Studies, College of Education
28 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan S7N 0X1
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

ABSTRACT

This study designed and prototyped an online semantization tool to learn and understand the *ginans*. The heritage of ginans of the Ismaili community comprises over 1,000 individual hymn-like poems that were composed and transmitted orally. The emotive melodies of the ginans continue to captivate even those members of the community who do not necessarily understand the language of the ginans. The language of the ginans is polyglottic as it borrows vocabulary from Indo-Aryan languages with loanwords from Perso-Arabic dialects. Thus, the language and the gnostic teachings of the ginans remain foreign to many of the English-speaking Ismaili community members in the diaspora.

This study is based on the premise that for the tradition and the teachings of ginans to be preserved in the Euro-American diaspora, the successive generations must learn and understand the vocabulary of the ginans. The process through which humans learn the vocabulary of a language is called *semantization*, which includes understanding various senses of the words. Access to lexicons such as dictionaries and glossaries remains critical for semantization. To this end, a sample ginan corpus was created and semantically analyzed in this study to develop an online ginan lexicon. This lexicon was then used to enrich the sample ginan texts with online glosses to facilitate semantization of the ginan vocabulary.

The design based-research methodology for prototyping the tool comprised two design iterations of analysis, design, and review. In the first iteration, the initial design of the prototype was based on a multidisciplinary literature review and an in-depth semantic analysis of the

available ginan materials. The initial design was then reviewed by community ginan experts and teachers to inform the next design iteration. In the second design iteration, the initial design of the prototype was further enhanced by adding features based on the needs of the community gathered by surveying a convenience sample of 515 community members from across the globe.

The analysis of the survey data revealed that over 90% of the survey participants preferred English materials for learning and understanding the language of the ginans. In addition, having access to ginan materials online was expressed as a dire need for the community to engage with the ginans. The development and dissemination of curriculum-based educational programs and supporting resources for the ginans emerged as urgent and unmet expectations of the community.

The study also confirmed that the wide availability of online ginan resources, such as the ones designed in this study, is highly desirable by the English-speaking community members who want to learn and understand the teachings of the ginans. However, these resources and tools are only part of the solution for fostering sustainable community engagement for the preservation of the ginans in the diaspora. To ensure that the tradition is carried forward by the future generations with compassion and understanding, the community institutions must make ginans an educational priority as well as ensure that educational resources for the ginans are developed and made readily available to the community through institutional communication channels.

ACKNOWLEDGMENTS

This dissertation is dedicated to the memory of the late Missionary Rai Dr. Abualy Aziz.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PROLOGUE: SITUATING SELF

While water takes on the shape of its container,

Without the water, the container has no purpose.

Similarly, *ginan* (knowledge) shapes the self

But without the *gur* (guide), the knowledge has no use.

<div align="right">Translated from Ghatpatni Dua attributed to Pir Sadardin</div>

I am a proud member of the Nizari Ismaili community. My community is a faith-based community and has historically been a minority community. Both my parents were born in India, where the community members were commonly referred to as the Ismaili Khojas. The religious path of the Khojas was known as *Satpanth* (True Path) in the pre-colonial Indian subcontinent. A vital heritage of the Satpanth Ismailis of India is the tradition of gnostic hymn-like poems known as the *ginans* (lit. gnosis). The authorship of the ginans is attributed to several preacher-saints who are known as *pirs* and *sayyids* in the community.

The late Dr. Abualy Aziz (1919–2008) was one of the most renowned and influential Ismaili missionaries with deep knowledge and expertise in the teachings of the ginans. He emigrated to Canada in 1985 with his family and settled in Burnaby, British Columbia. I had the privilege of being his student and it was under his guidance that I initially started organizing the ginans some twenty years ago. In one of my last personal meetings with him, he encouraged me to continue the work of preserving the ginans.

I studied systems science at Simon Fraser University and graduated with B.Sc. in 1998. Upon graduation, I worked as an information technology consultant in Vancouver with PwC and IBM Canada. In my spare time, I started gathering and organizing ginan materials for posterity. After struggling to do so for a few years, it was clear to me that my information management and technology skills alone were not sufficient for the task of preserving the ginans. Consequently, I enrolled in the Master of Library and Information Studies at the University of British Columbia (UBC) in 2006 to learn about managing and organizing collections for access and preservation. By the time I graduated from UBC in 2009, my professional outlook had been transformed.

I found myself drawn to academic librarianship due to the shared personal and professional values of service, equity, learning, and a deep conviction for facilitating knowledge accessibility and preservation. I joined the University of Saskatchewan (USask) in 2009 as an information technology librarian and secured my tenure in 2014. Being at USask allowed me to work on ginans as part of my formal program of research. As part of this research, I developed Ginan Central, which now serves as a global portal for the long-term access and preservation of the ginans to foster research and education.[1]

Working on the ginans over these years, I have come to realize that preserving ginans requires an interdisciplinary approach and community engagement. This research is the beginning of my journey to develop such an approach to safeguard the tradition of ginans.

Karim Tharani

Saskatoon, 2021

---

[1] Ginan Central is hosted at the University of Saskatchewan Library and can be accessed at https://ginans.usask.ca.

CHAPTER 1: INTRODUCTION

An essential prerequisite for a community to pass on its traditional knowledge is the acceptance and learning of this knowledge by its successive generations. This acceptance is effective when based on genuine understanding and compassion nurtured through contextual teaching and transmission. The cultural contexts and experiences of ethnocultural community youth born and raised in the Euro-American[2] diaspora are very different from their elders who had to leave their native lands and languages behind. Thus, these young members in the diaspora often find themselves having to navigate the demands of community traditions and pressures of popular culture on their own. As information technology permeates our everyday lives, it is natural for the youth of ethnocultural communities living in the diaspora to want to engage with their traditions and heritage *on-line* and *on-demand*.

Unfortunately, the historical discourse of modernity often pits non-Western traditions against technological progress in the West. Moreover, information technology is also often perceived as instrument of appropriation and colonization in certain ethnocultural communities. However, perhaps due to its minority status as a religious community without any permanent homeland, adapting to changing times and technologies has been a constant and consistent characteristic of the Ismaili community in general.

Context – The Nizari Ismaili Community and the Corpus of Ginans

Oral traditions have long served as vehicles of veneration and education in Muslim societies, tracing their origins back to the revelation of the Quran itself. Over time and

---

[2] The term "Euro-American" (as opposed to Western) more accurately conveys civilizational affinities rather than geographical boundaries and also includes countries that are not geographically "Western" but remain "European" in terms of their civilization, such as Australia, New Zealand, etc. (Alibhai, 2020).

geography, these traditions have found diverse expressions accommodating local cultures and languages. At times these traditions have taken melodic forms by leveraging the power of poetry and music to bring faith, knowledge, devotion, and worship into ecstatic harmony. An exemplar of such melodic expressions is the heritage of the *ginans* of the Ismaili Muslim community.

The word *ginan*, a derivative of the Sanskrit term *jnan*, means knowledge or gnosis. In the context of the Nizari Ismaili community, the term is used for the community's collection of religious hymn-like poems. The ginans were originally composed and transmitted by the early preachers to spread the Satpanth Ismaili faith in South Asia. Thus, the ginans once served as scriptural texts that guided the normative understanding of the faith for the Satpanthi Ismailis in South Asia. There are over 1000 individual ginans of varying lengths and languages that constitute the religious heritage of ginans.

While the ginans are the heritage of the Satpanthi Ismailis of South Asia, the global Ismaili community is a culturally diverse community living in over 25 countries around the world with various religio-cultural traditions and heritage, including various oral traditions, such as qasida, munqabat, munajat, and geet.[3] Today, the ginans continue to be cherished in the diaspora, albeit as one of many of the community's diverse devotional and cultural traditions. In the context of this study, the term *community* is used for those members of the Nizari Ismaili community in the diaspora who recognize and revere ginans as part of their religio-cultural heritage.

Since ginans were transmitted orally, the language of ginans is also oral and without an established writing system or script of its own. The community devised a script called *Khojki* to

---

[3] For more information on the Ismaili community, please refer to the official community website available at https://the.ismaili/global/about-us/the-ismaili-community.

transcribe the ginan texts in manuscripts (Asani, 2002). Tracing its roots to mercantile communities of India, Khojki is not a language but a script that was utilized as a shorthand for jotting down accounts.

With the advent of printing technology, the community founded the Khoja Press of India in 1903 to create an official canon of the authorized texts of the ginans in Khojki. The press used specialized German-made fonts for the Khojki script for printing the ginan corpus (Asani, 2011). Later, the community embraced Gujarati as its *lingua franca*, and the ginan texts were converted and printed in Gujarati for the community use. As the community members migrated away from the Indian subcontinent to Western countries in the 1970s, the ginan texts were printed using the Latin or English script. Consequently, the ginan corpus today includes texts in various scripts, including Khojki, Gujarati, Urdu, and English (Latin).

While the instinctive adaptability of the community in the face of print technology has been instrumental in preserving and transporting the corpus of ginans over time and geography, the teachings of the ginans remain foreign to many community members. This estrangement from the ginans is particularly true for the English-speaking community members in the diaspora. Consistent with this historical adaptability of the community in the face of changing times, this research puts forward an approach to safeguard the Ismaili community's tradition of *ginans* with deliberate and purposeful use of information technology. The study prototypes an online tool that incorporates the Western best practices of language learning with the traditional methods and materials of teaching and transmitting ginans.

Purpose of the Study

This study is based on the premise that for the tradition and teachings of the ginans to continue in the Euro-American societies, including Canada, the successive generations must learn and understand the vocabulary of the ginans. The vocabulary of the ginans can be categorized as *polyglottic* (or mixed language) as it borrows vocabulary primarily from Indo-Aryan languages (e.g., Gujarati, Hindi, Sanskrit, etc.) with loanwords from Perso-Arabic dialects (Shackle & Moir, 2000). With the language of the ginans being oral and mixed, it remains foreign to many of the English-speaking community members living in the Euro-American diaspora. This study frames the challenge of learning and understanding the ginans as equivalent to learning a foreign language by learners whose primary language is English. The purpose of this study is to prototype an online ginan learning tool that balances the current needs and expectations of the English-speaking community learners with the established practices of teaching ginans in the community.

Research Questions

To effectively research, design, and validate an online ginan learning tool, the study focused on the following overarching research question:

RQ: *How can an online tool be designed to facilitate the learning and understanding of the ginans by English-speaking learners?*

Further analysis of the different aspects of this broad research question led to the formulation of three sub-questions. Each sub-question serves as a building block of the overall framework for this research as depicted in the figure below (Figure 1.1).

RQ1: How can the existing ginan corpus be curated to support online language learning from a digital humanities perspective?

RQ2: What are the needs of the community members interested in learning and understanding the ginans?

RQ3: In what ways can information technology be useful in learning the language of ginans?

Research Objectives

The objectives of this research were to:

1. Synthesize, analyze, and digitally curate relevant existing materials to support online ginan learning.

2. Conduct a needs assessment of community members interested in learning and understanding the language of the ginans.

3. Combine the traditional teaching practices of the community with the contemporary learning theories and best practices.

4. Design, prototype, and validate an online tool to learn the vocabulary of the ginans.
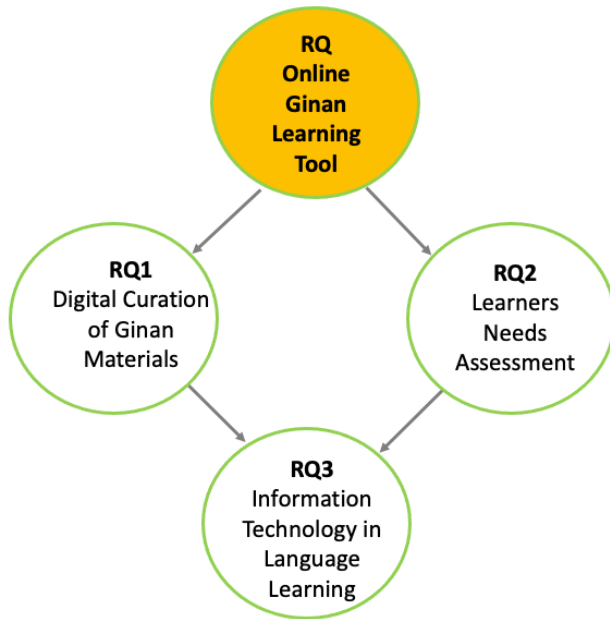
Figure 1.1. Relationship between Research Questions

Significance of the Study

A language remains alive if there are people who continue to use the language in their everyday lives. For a language to remain relevant to people and to attract new learners, it must also have a growing base of popular, intellectual, and instructional materials in a variety of formats (text, audio, video, web, etc.) that language users can consume and contribute to in their professional and personal lives. With the proliferation of information technology, more and more materials, especially language resources, are being made available online and in multimedia formats. As a result, more learners are choosing to use self-help software and online tools to learn foreign languages. There are commercial organizations that specialize in creating self-study multimedia resources to teach native or foreign languages.

Over time, however, some languages fail to attract new learners, and this failure causes them to become less popular or even forgotten by the general public. Due to the limited interest

in such languages, developing self-help tools is no longer profitable for commercial companies. In such cases, the responsibility of safeguarding endangered languages is often shouldered by researchers and scholars. This study makes tangible contributions to identifying elements of an enabling online environment for learning oral and mixed languages that are on the verge of being forgotten.

The resulting design and prototype provide a concrete template for leveraging information technology to find common ground amongst competing interests, needs, philosophies, and practices for passing on traditional knowledge to successive generations. This study will be helpful for the Ismaili community stakeholders in developing strategies and programs to make the ginans more accessible and relevant to the younger generations living in the diaspora.

The study advances the existing body knowledge on language preservation in a religio-cultural context. It is the first study of its kind to use information technology tools in the learning and understanding of the ginan tradition. This study interlaces several disciplinary areas (including computer science, digital humanities, linguistics, and educational technology) to identify and leverage disciplinary best practices that can be applied to the transmission and preservation of non-Western languages and traditions.

In a broader context, this research benefits other ethnocultural and Indigenous communities with rich heritage of traditional knowledge by enabling them to make informed decisions on how best to transmit and teach (and thereby safeguard) their oral traditions and languages in modern times. An equally important contribution of this study is the harmonious application of the contemporary Western best practices of language learning in a non-Western

educational setting.

Organization of the Study

This research explores how information technology helps combine the contemporary best practices of learning a foreign language with the traditional teaching and transmission of ginans. It uses design-based research (DBR) to prototype an online ginan learning tool. With DBR, learning interventions and tools are collaboratively designed and evaluated by researchers and practitioners (Melgosa, 2018). In addition, the research utilizes the educational research and development (ER&D) method for prototyping as defined and outlined by Borg and Gall (2003). The organization of this study is as follows:

Chapter One provides contextual information, the purpose of the study, research questions and objectives, significance of the study, and organization of the study, followed by the definition of pertinent concepts and terms used in this study.

Chapter Two is a broad multidisciplinary literature review on the topics that are pertinent to the purpose of this research. These topics include corpus processing, semantic analysis, learning foreign languages, as well as digital humanities. The chapter closes with a summary of the literature review, which is followed by remarks on situating the ginans in the context of the literature review.

Chapter Three outlines how the design-based research approach and the education research and development method for prototyping are paired together for designing the online ginan learning tool. It also outlines the research design and the methods used to produce the prototype.

Chapter Four reports the results of various analyses undertaken as part of prototyping the online ginan learning tool, including semantic analysis, design iterations, expert reviews analysis, learners' needs analysis, followed by a discussion on the prototype design and development.

Chapter Five includes a summary of activities, conclusions, research outcomes, limitations, followed by recommendations for future studies.

Definition of Terms

The research draws its terminology from multiple disciplines, including computer science, corpus linguistics, library information studies, and digital humanities. In addition, specific concepts and terminology of the Ismaili community are also embedded in this document. The following definitions of pertinent terms are included for reference.

Concordance	An alphabetical list of words in a text or group of texts, with information about where in the text each word occurs and how often it occurs. The sentences in which each word occurs are often given (Collins Online English Dictionary, 2020).
Corpus	A large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria (Bowker & Pearson, 2002).
Diacritic	A sign placed above or below a character or letter to indicate that it has a different phonetic value, is stressed, or

for some other reason (Collins Online English Dictionary, 2020).

Diaspora	People who come from a particular nation, or whose ancestors came from it, but who now live in many different parts of the world are sometimes referred to as the diaspora (Collins Online English Dictionary, 2020).
Ginans	Derived from the Sanskrit word *jnan* meaning meditative or contemplative knowledge. It is a general term used for the corpus of the indigenous devotional literature of the Nizārī Ismaili Khojas and some related groups of South Asian origin. Composed in a number of Indian languages and dialects of Sind, Panjab, and Gujarat, these hymn-like poems, which continued to be composed and revised until the early decades of the 20th century. (Daftary, 2011).
Ismailis	The Shia Imami Ismaili Muslims, generally known as the Ismailis, belong to the Shia branch of Islam. The Shia form one of the two major branches of Islam, the Sunni being the other. The Ismailis live in over 25 different countries, mainly in Central and South Asia, Africa, the Middle East, Europe, North America and Australia, and number approximately 20 million. The Ismailis continue to believe in the line of Imamat in hereditary succession continuing

from Ismail to His Highness the Aga Khan, who is their present, 49th Imam in direct lineal descent from Prophet Muhammad (TheIsmailis.org, 2020).

Khojas   Name commonly used in reference to the Nizārī Ismailis of South Asian origin (Daftary, 2011).

Khojki   Name of a script developed within the Nizārī Khoja community of South Asia (Daftary, 2011).

Lexicon   The body of words in a language or the words associated with a particular subject (Collins Online English Dictionary, 2020).

Mixed language   A language that arises among a bilingual group combining aspects of two or more languages but not clearly deriving primarily from any single language (Wikipedia, 2020).

Polyglottic   Written in, composed of, or containing many languages. (Collins Online English Dictionary, 2020).

Polysemous   A Greek term that is used for words that have many *word senses* (Jurafsky & Martin, 2019).

Proof-of-Concept (PoC) The stage during the development of a product when it is established that the product will function as intended (Collins Online English Dictionary, 2020).

Prototype   A preliminary design of the educational product that is to be field-tested (Borg & Gall, 2003).

Romanization	Romanization, in linguistics, is the conversion of writing from a different writing system to the Roman (Latin) script, or a system for doing so. Methods of romanization include transliteration, for representing written text, and transcription, for representing the spoken word, and combinations of both (Wikipedia, 2020).
Satpanth	"True Path." Technical Indian designation for the indigenous Ismaili tradition of the Nizari Khojas of South Asia (Daftary, 2011).
Semantization	A continuing process of getting acquainted with verbal forms in their polysemous diversity within varying contexts (Beheydt, 1987).
Transcription	Transcription in the linguistic sense is the systematic representation of language in written form. The source can either be utterances (speech or sign language) or preexisting text in another writing system (Wikipedia, 2020).
Transliteration	Transliteration is a type of conversion of a text from one script to another… [and] is not primarily concerned with representing the sounds of the original but rather with representing the characters, ideally accurately and unambiguously (Wikipedia, 2020).

Wireframe	A visual representation of the structure of a web page (Collins Online English Dictionary, 2020).
Word sense	A discrete representation of one aspect of the meaning of a word (Jurafsky & Martin, 2019).
Word Sense Disambiguation (WSD)	The task of determining which sense of a word is being used in a particular context (Jurafsky & Martin, 2019).

CHAPTER 2: LITERATURE REVIEW

When we study language, we are approaching what some might call the "human essence," the distinct qualities of mind that are, so far as we know, unique to man.

— Noam Chomsky, *Language and Mind*, 2006.

This chapter undertakes a multidisciplinary literature review of several topics that are pertinent to the purpose of this research. These topics include corpus processing, semantic analysis, foreign language learning, as well as digital humanities. A summary of how these topics inform the purpose of this research to design an online tool for the ginans is provided at the end of the chapter.

Preserving Endangered Mixed Language

A *mixed* language is formed through the phenomenon of "language contact," whereby people speaking multiple languages interact with each other at the same place and time (Bakker & Mous, 1994). For this reason, mixed languages are also referred to as *contact* languages. Initially, it was believed that Michif of the Métis people was the only known mixed language of the world, but other mixed languages have also been identified since then (Bakker, 2017). But even today, only a handful of over 7,000 documented languages are classified as contact languages (Lewis, Simons, & Fenning, 2020). Some of these mixed languages, like Michif, are also endangered and in dire need of preservation to remain alive (Barkwell, 2017; Petten, 2006).

There are three primary strategies to preserve endangered languages. These are: 1)

creating and preserving the corpus of spoken and written language materials, 2) semantically analyzing the corpus to document the linguistic and grammatical aspects of the language, and 3) developing pedagogical materials to transmit the language to successive generations of learners (Barkwell, 2017; Yang & Rau, 2005). These interdependent strategies ensure that a language and the knowledge it encapsulates remain preserved and accessible to the current and future language learners.

In the literature on language learning, the terms L1 and L2 are commonly used to refer to the first and second languages of the learners. The rationale is that a learner's native language (L1) is the one that they first learn as a child. Any additional language that they learn during their life is regarded as their second (L2) or additional language. Access to lexicons such as dictionaries and glossaries remains critical for L2 learners. A dictionary provides grammatical information, definitions, synonyms, antonyms, as well as examples of usage; whereas a glossary is typically developed for learning special or subject-based vocabulary. The study and creation of corpus-based lexicons is the primary focus of Corpus Linguistics (CL), which is a subfield of applied linguistics. The focus on corpus-based lexicons in CL is steeped in the philosophy to privilege real-life textual samples of language over theories and intellectual constructs for the study of languages (Bowker & Pearson, 2002).

Corpus Processing for Lexicons

CL traces its origin to the 1950s when samples of textual materials were used in the study of languages. Back then, these samples "might have been simple collections of written or transcribed texts and thus not representative… and used primarily for the study of distinguishing features in phonetics" (McEnery, Xiao, & Tono, 2006, p. 3). This approach was perhaps the

17

reason why CL initially came under heavy criticism for 'skewedness' by traditional language researchers as it emerged as a full-fledged discipline within applied linguistics (Libben, 1990). Since then, however, with better sampling and automated processes, CL has evolved as a robust discipline within linguistics, which dedicated to the study of language based on examining a collection of authentic and representative textual materials (or corpora) using real-life samples and empirical methods.

Creating a Corpus

The contextual and contemporary definition of the term *corpus* (pl., *corpora*) in CL today is that it is "a large collection of *authentic* texts that have been gathered in *electronic* form according to a specific set of *criteria*" (Bowker & Pearson, 2002, p. 9). Given that people study languages for a variety of reasons, there could be many different types of corpora such as general or special purpose corpus, monolingual or multilingual corpus, etc. The type of corpus to be created depends on the purpose of the study and the language under study. For example, a special purpose corpus may be created for learning subject-specific vocabulary and the semantic relationships among the words. Regardless of what type of corpus is used in the study of a language, a corpus is not intended to be a replacement for instructors or subject matter experts.

The use of a corpus is not just limited to vocabulary learning but is also common in natural language processing (NLP). In the context of NLP, corpus-based processing is often called *computational linguistics* to emphasize the focus on machine learning and artificial intelligence. The history of NLP is often traced to Alan Turing for devising a test (now known as the "Turing Test") to gauge a computer's ability to exhibit intelligence. With the use of corpus for learning languages by humans and machines, corpus remains at the heart of enhancing human

understanding and artificial intelligence in language learning.

A *parallel corpus* is used when working with a mixed language. A parallel corpus consists of a source corpus and its translation in one or more languages. The source and translated corpus are aligned with each other, often automatically, by making use of corpus markups. This alignment or links are created between corresponding sections, paragraphs, or sentences in works of prose, and cantos or verses in works of poetry. The granularity of alignment depends on the compatibility between source and translated corpora within a parallel corpus (Bowker & Pearson, 2002).

From an academic perspective, mixed languages provide "rare insights in understanding sociological, psychological, and cognitive factors that allow humans to take two or more typologically unrelated fully functional languages, split them, and create a new entirely functional language based on different linguistic components with minimal blending" (Schreier, 2015, p. 480). An important aspect of mixed languages is the notion of *loanwords*–words adopted from a foreign language into a local language. Before the use of computers in CL, scholars and practitioners relied primarily on their intuition to identify loanwords to assess the influence of one or more languages on a given mixed language. With the increase in the use of specialized computer applications in CL, identification of candidate loanwords in mixed languages has also become much more efficient, thorough, and quantifiable, as Bowker (2018) explains:

> Consider that a corpus is a text file. It could be made up of tens, hundreds or thousands of documents and may run to hundreds of thousands or even millions of words. Trying to count the number of words, or the number of times each word occurs, would be a time-

consuming, labor-intensive and error-prone process if it were done manually. However, this type of work is easily accomplished by a computer, and corpus analysis software can be used to calculate several different measures of frequency, including raw frequency counts (e.g., word lists), measures of disproportionate frequency (e.g., keyness), and measures of relative frequency (e.g., collocations). (p. 361)

The Corpus of Ginans

As was noted earlier, the efforts to canonize the texts of ginans by the Ismaili community can be traced back to the late 19th century. Lalji Devraj (1842-1930) of India is credited for publishing the initial canon of authorized texts of ginans in Khojki as early as 1903 (Asani, 2011, p. 25). In the first quarter of the twentieth century, the responsibility to publish the official texts of the ginans moved away from individuals to community institutions – starting in 1922 with the Recreation Club Institute in India and followed by the national branches of the Ismailia Association and Ismaili Tariqah and Religious Education Board (ITREB). [4]

In 1977, with the establishment of the Institute of Ismaili Studies (IIS)[5] in England, the mandate of developing curriculum and materials for the primary[6] and secondary[7] religious education gradually shifted from national ITREBs to IIS. As a result of this shift, the responsibility of ensuring that the community has access to authorized and authentic texts and other educational ginan materials has essentially been ignored, as Asani (2021) notes:

---

[4] The Ismaili Tariqah and Religious Education Board (ITREB) is a global volunteer-based network of national and regional community committees responsible for administration of religious education of the Ismaili community.
[5] The Institute of Ismaili Studies (IIS) is an academic institution established in 1977 with the mandate to promote historical and contemporary scholarship and learning on Muslim societies and cultures and their relationship with other societies and faiths (www.iis.ac.uk).
[6] More information on the IIS Primary Curriculum is available at https://www.iis.ac.uk/curriculum.
[7] More information on the IIS Secondary Curriculum is available at https://www.iis.ac.uk/graduate-studies/iis-secondary-curriculum.

[T]he divergence – particularly evident in recent decades – between official Ismaili institutional perspective on the *gināns* and popular engagement with them by Ismailis in their personal capacities. Anxieties about perceptions that other Muslims may have of the ginans, in particular their vernacular Indic character, have been the primary concerns to Ismaili institutions. These concerns have led to a marked de-emphasis of the semantic dimension of the *gināns* in the contemporary articulation of official Ismaili doctrine in favor of a Quranic one. Instead, there is an increased focus on the performative aspects of the *gināns*, and their ritualization as a form of Ismaili "devotional literature," thus reframing them within the context of Ismaili literary traditions in Arabic and Persian. (p. 50)

Consequently, local community missionaries, also referred to as Al-Waez in the community, took it upon themselves to preserve and propagate the scriptural status of the ginans. Most notably, the efforts of the late Missionary Abualy Aziz in this regard have been significant as mentioned by Alibhai (2014):

Al-Waez Abualy made the study and teaching of Ginans the centerpiece of his lifework as preacher and scholar. He painstakingly built a huge library of Ginan texts, some of which are in the form of manuscripts written in Gujarati or Khojki scripts, and cared for them with great devotion and sacrifice, bringing them with him first from India to Africa, then from Africa to Canada. (p. 48)

Other elders and preachers in the community have also produced English texts and translations of the ginans to help teach and transmit the tradition and its meaning to the younger generations who may not be accustomed to the Indic languages and cultural motifs used in the ginans. The

efforts of Kamaluddin Ali Muhammad and his wife Zarina Kamaluddin, notes Virani (2015),

have made ginans more accessible and comprehensible to those not familiar with Gujarati and

Khojki scripts:

> Al-Waʿiẓ Kamaluddin Ali Muhammad and al-Waʿiẓa Zarina Kamaluddin have made
>
> Herculean efforts to study and translate this literature. Their work has added
>
> tremendously to our knowledge of not only the Ginans, but of mediaeval South Asian
>
> verse in general, for the Ginanic symbols and vocabulary draw on the rich universe of
>
> mystical meaning that had become the common inheritance of Sufis, yogis, sadhus,
>
> bhaktas and sants. All students of this field and lovers of the Ginans are indebted to their
>
> endeavours. (p. viii)

It must be noted, however, that the materials authored by Kamaluddin Ali Muhammad have

often attracted criticism for "Islamization" of the ginan texts such that the "interpretation of

teachings of the ginans – their metaphors, their symbols, [and] their allegories – were mostly

now limited to Quranic framework" (Asani, 2021, p. 39).

A direct consequence of the dispersed and independent production of ginan materials has

been the use of variant romanization conventions in transliterating the ginans from Khojki and

Gujarati to the Latin script. As a result, the available ginan texts come in a variety of

romanization conventions that range from very simple to highly sophisticated. These

incompatible romanization conventions make it difficult to create an integrated electronic corpus

for teaching and research.

A positive development in the evolution of the ginan corpus has been the "modern

scholarly study of Ginans [which] is more than fifty years old, pioneered by the great Russian

scholar of Ismailism, W. Ivanow" (Alibhai, 2014, p. 48). In an edited volume on Satpanth

Ismailism published by Brill, Ivanow included a few specimens of the ginans (1948). By virtue

of this publication, the ginans debuted in academic literature. Since then, there have been over

twenty dissertations submitted on ginans in Euro-American universities, including Harvard,

Columbia, and McGill.[8] In 2020, the IIS also established the South Asian Studies (SAS) unit

with the mandate to "contribute to current academic debates as they relate to Islam and Muslims

in South Asia and to further scholarly understanding of Satpanth history, literature, heritage and

identity in order to promote critical thinking in the field of South Asian Ismaili Studies."[9] While

the mandate falls short of mentioning the ginans, it is hoped that SAS will make creation of an

online corpus of ginans for research and education a part of its academic strategy.

Corpus Processing Methods and Techniques

Hunston (2002) identifies *frequency* and *collocation* as the building blocks for creating

corpus-based lexicons such as concordances, glossaries, and dictionaries. In CL, the frequency is

a raw count of occurrence of linguistic features (such as words, word types, phrases, etc.) in a

corpus. Frequencies are commonly used to generate wordlists that are ranked by the frequency of

words while analyzing corpora. The collocation is an automated process for identifying words

and word patterns that typically occur in each other's vicinity. Collocations are used to generate

concordances, also referred to as keyword-in-context (KWIC). A concordance is an alphabetical

display of a given word or phrase in a corpus together with its immediate contextual occurrence.

Corpus *markup* and *annotation* are two common processing techniques that facilitate

---

[8] A comprehensive bibliography of Ginan Studies is available at https://ginans.usask.ca/studies.
[9] The South Asian Studies (SAS unit became operational within the Department of Academic Research and Publications (DARP) upon the hiring of its inaugural Academic Lead in January 2020. For more information, see https://www.iis.ac.uk/content/south-asian-studies.

parsing and tagging of corpus to identify lexical and grammatical features of a language, including word sense disambiguation, and part-of-speech (Martinez, 2010). Corpus markup is the process of "automatic insertion of marks, known as tags, to make the appearance and structure of text explicit" (Bowker & Pearson, 2002, p. 231). The nature and number of tags in text processing are driven by the purpose of the research for which the corpus is assembled. In general, however, corpus markup allows an interrogation of corpus to investigate research questions. Closely related to the technique of corpus markup is the technique called corpus annotation. Corpus annotation is the process of appending "interpretive, linguistic information to an electronic corpus of spoken and/or written language data" (Leech, 1997, p. 2).

Creating Corpus-Based Lexicons

A lexicon can be defined as an alphabetical list of terms in one or more languages organized to capture pertinent information of interest in a systematic manner. The amount of information contained in a lexicon can vary depending on its scope and purpose. At one end of the spectrum is a basic lexicon such as a glossary, which lists terms and their equivalents in one or more foreign languages. At the other end of the spectrum are richly detailed resources, such as a dictionary, containing definitions, examples of usage, synonyms, and related terms. However, practical lexicons fall somewhere in the middle of the spectrum and would typically contain definitions, equivalents and, usage information. The following process is used in CL as a best practice for creating a lexicon (McEnery, Xiao, & Tono, 2006):

1. *Corpus gathering and processing:* The purpose of a lexicon must guide what materials are selected to build a representative corpus. Markups and annotations are then inserted in the corpus for analyses and identification of candidate terms.

24

2. *Identifying candidate terms:* A candidate term is a term that has been identified, either manually or through means of automated process, as one with the potential to qualify for inclusion in the final lexicon. It is a basic meaning-bearing lexical unit of a language and may consist of one or more words.

3. *Selecting content terms:* The content terms are words (nouns, verbs, adjectives, and adverbs) that have a clearly defined meaning. They are chosen to be included in the lexicon from the pool of candidate terms either manually or via an automated algorithm.

4. *Developing an online semantic profile:* A semantic profile describes various linguistic characteristics and relationships of the content terms, such as parts of speech and synonyms.

In addition to documenting linguistic and grammatical features, semantic analysis is particularly useful in ascertaining the semantic relationships of the words in a language and to disambiguate various senses of the words. For L2 learners to be successful, Beheydt notes that "it is essential that the learner be provided with a number of concrete representative usages of each word as a basis for the correct semantization of a word" (1987, p. 61).

Semantic Analysis

The second strategy for preserving a language is to semantically analyze the language corpus for posterity. Semantic analysis is the process of "systematically comparing and contrasting related words, and summarizing the similarities and contrasts in the most economical way" (Goddard, 2011, p. 51). The use of semantic analysis is not limited to linguistics but is widely utilized in humanities and sciences. The approach to semantic analysis, however, may

vary based on disciplinary traditions as well as research goals.

In his seminal study of ethical terminology in the Quran, Izutsu defines his approach to semantic analysis as "an analytic study of the key words of a language with a view to arriving eventually at a conceptual grasp of the Weltanschauung or worldview of the people who use that language as a tool not only of speaking and thinking, but, more important still, of conceptualizing and interpreting the world that surrounds them" (Izutsu, 1964, p. 11).

Approaches to Semantic Analysis

In the context of this study, the use of semantic analysis can be placed on a continuum ranging from unstructured to structured approaches. The unstructured approach, also known as the *bag-of-words* approach, represents one end of the spectrum where documents are treated as random collections of words. In this approach, the order and relationship of words to each other are completely ignored. This approach is commonly used by commercial web search engines to facilitate keyword searches (Freitas, Curry, Oliveira, & O'Riain, 2012). Despite its scalability, the unstructured approach fails "to account for differences in meaning as a result of differences in word order or syntactic structure (e.g., 'cats climb trees' vs. 'trees climb cats')." (Tai, Socher, & Manning, 2015, p. 1556). This failure is the primary reason why the search engines like Google continue to generate voluminous and irrelevant search results for keyword searches.

Lying in the middle of the spectrum is the semi-structured approach to semantic analysis, whereby meanings of texts are assigned representative labels and categories. This approach is used in libraries, for instance, to develop and assign subject headings (e.g., Library of Congress Subject Headings, Dewey Decimal Classification, etc.) to facilitate classification, search, and retrieval of relevant materials. Librarians make use of semantic analysis to make "replicable and

valid inferences from texts (or other meaningful matter) to the contexts of their use"

(Krippendorff, 2004, p. 18). Topic modelling is another application of the semi-structured

approach to semantic analysis to "connect words with similar meanings and distinguish between

uses of words with multiple meanings" (Sharma & Sharma, 2017, p. 84).

On the far end of the spectrum is the structured approach to semantic analysis, which is

highly prevalent in computational linguistics. Computational linguistics combines resources from

linguistics and computer science to develop computational models to understand how human

language works. The purpose of such scientific approaches to the study of language is to create

practical tools such as text parsers, spelling and grammar checking modules, speech recognition,

machine translation, text mining, and others (Mishra, 2018).

The Use of Computers in Semantic Analysis

The creation of Index Thomisticus in 1949 by Roberto Busa, a priest and philologist, is

often cited as the beginning of *humanities computing* – i.e., the use of information technology

and computing in humanities. Busa began with the question of "inwardness" in the writings of

St. Thomas Aquinas as part of his doctoral dissertation. He asserted that the only dependable

way to understand how Aquinas perceived *inwardness* was through finding all the occurrences of

the term in the corpus of his writings. This inquiry led Busa to undertake the creation of Index

Thomisticus, a concordance of the vocabulary of St. Thomas Aquinas (Busa, 1980). Given this

monumental task, Busa was determined to automate the process and collaborated with computer

technologists at IBM. Reflecting on his experience and the nature of the nascent human-machine

partnership, Busa (1980) notes:

Linguistic research must tend to be based on full inventories of words at least of large

corpuses. Semantics is the description of the kingdom of the creative freedom of spirit: in man it is his soul which talks. Consequently scientific description of how we talk can be nothing but probabilistic…This is the reason why the use of computers in linguistics demands a lot of dedication and hard work. Without them [humans], computers would only produce "in real time" monuments of waste. (pp. 89-90)

Perhaps because of Busa's initiative, CL as a discipline was an early adopter of computers for semantic analysis during the personal computing era back in the 1990s. The most significant contribution of CL to the study of languages has been the machine-readability of texts via corpus processing tools and techniques to facilitate semantic analysis.

Following Busa's successful use of computing, the US government further formalized the use of information technology in humanities by funding the Machine Translation Project during the Cold War in 1966. This project was based on a proposal by Warren Weaver, an American scientist, who claimed that machines can be programmed to translate languages much faster and more reliably than humans (Weaver, 1949). In stark contrast to Busa's view, however, Weaver saw semantic analysis as a mere mechanical process "concerned with the interpretation of meaning by the receiver, as compared with the intended meaning of the sender" that could be automated without human intervention (p. 11). The project ended abruptly when it was determined that it was not economically feasible to replace the work of human translators with computers. The experiment, nevertheless, propelled scholars in the sciences to collaborate with their peers in humanities to develop computational models and applications for the study of languages through semantic analysis.

Semantic Analysis and the Web

Tracing the history of humanities computing since the creation of Index Thomisticus in 1949, Professor Willard McCarty, a leading digital humanist, regards the onset of the World Wide Web (or the Web) as a "flood" that swept scholars and practitioners away from their disciplinary bases (McCarty, 2015). Continuing with his metaphor of a flood (or diluvia), McCarty partitions the history of computing in humanities into three phases, namely: Antediluvian, Diluvian, and Postdiluvian. In this scheme, McCarty asserts that in the Antediluvian phase, starting from 1949 to 1991 (i.e., the period before the Web was invented), linguistic analyses dominated the use of digital technology. In the Diluvian phase, which centred on the invention of the Web in the 1990s and its proliferation until 2005, the emphasis moved away from linguistic analyses to digitization and multimedia to create digital collections and libraries. It was not until after 2005 in the current Postdiluvian phase, according to McCarty, that scholars and practitioners have found their bearings and are beginning to develop disciplinary self-awareness in directing the use of digital technology in humanities rather than being directed by it.

Thus, in today's Postdiluvian phase, the Web serves as a global computing platform and facilitates interdisciplinary research. The Web continues to facilitate "both the development of software tools (for information/document retrieval, document summarization, sentiment analysis, named entity extraction, corpus analysis and machine translation) and the development and refinement of linguistic theories via algorithmic models" (Bateman, McDonald, Hiippala, Couto-Vale, & Costezki, 2017, p. 2).

Evolution of the Web

Since its invention in the early 1990s by Tim Berners-Lee, the Web has served as a portal

to connect disparate information resources through hyperlinks. The uptake of this new hypertext media was immediate, and it was very consistent with McCarty's metaphor of a flood, as the public started making "material available on the Web for all conceivable reasons—to communicate, educate, persuade, promote, defraud, and delude, to name just a few of the possibilities." (Furuta, Shipman III, Marshall, Brenner, & Hsieh, 1997, p. 167). Subsequently, the Web became a global interactive and social platform for users to share all aspects of their lives. Within the first decade of its inception, the Web had over 29 million registered websites and over 500 million users (Internet Live Stats, 2017). With such an uptake, societal needs and expectations also evolved around the need to retrieve and access information over the Web. Thus, technologies supporting the Web platform increased in speed, power, and sophistication to enable dynamic publishing of data from personal and organizational databases as well as multimedia content such as music and videos.

The Google Effect. In 2001, Google made its debut as an online service that offered free access to information over the Web through its search engine. While Google made this service free for the public, it charged commercial organizations for advertising on its website. The arrival of Google ushered the Web into an age of unprecedented commercialization that ran counter to the tenets of transparency and open access of information that Berners-Lee had hoped for (Berners-Lee, 2012). As Tharani (2019) points out,

> [Very] few members of the public today pause to wonder why searching for a book using Google's search engine never leads them to a library. Even more disconcerting is the fact that the rules and algorithms used by Google to share or suppress information on the Web remain obscure and proprietary. (p. 8)

Within a decade of its debut, Google emerged as a self-appointed information gatekeeper for the Web to maximize its commercial interests and profitability.

The Semantic Web. The awe and excitement surrounding the birth of the Web mobilized people into exploring new possibilities of the Web. In this state of exploration and imagination, very few – if any at all – anticipated the risk to the fair and open accessibility of public information due to the commercialization of the Web. Having not anticipated such a fate for the Web initially, Berners-Lee revealed his new vision for the Semantic *Web* (Berners-Lee, Hendler, & Lassila, 2001).

At its core, the Semantic Web presents an alternative model for publishing, searching, and accessing information over the Web without commercial gatekeepers or mediators. The Semantic Web is designed to bring rigour and structure to transform the current Web of *documents* into the Web of *data* (Berners-Lee, Hendler, & Lassila, 2001). The Semantic Web does so by making the *semantics* of contents on the Web discernible to machines via knowledge *ontologies* and intelligent *agents*. In the context of the Semantic Web, *ontologies* are statements that define the relations between concepts and specify logical rules for reasoning about them; whereas an *agent* is a piece of software that runs without direct human control to accomplish goals provided by a user that requires them to collect, filter, and process information found on the Web, sometimes with the help of other agents (Berners-Lee, Hendler, & Lassila, 2001).

Language Learning and Semantization

The final strategy for preserving languages is the transmission and teaching of the language by successive generations. The process through which humans learn words of a language and their meanings is called *semantization*. An important aspect of vocabulary

semantization is recognizing that words are *polysemous* by nature, that is they are used in multiple senses. A word sense "is a discrete representation of one aspect of the meaning of a word" (Jurafsky & Martin, 2019, p. 354). Therefore, having access and exposure to language corpus is an important prerequisite for language learning. As Poole asserts, "students exposed to authentic usages of vocabulary in multiple meaningful contexts may more efficiently semanticize and more completely acquire vocabulary than learners who are not exposed to the varied usages of lexical items" (2011, p. 3). For the purpose of this study, semantization refers to the ability of learners to recognize and understand the meaning of words in various contexts.

Semantization in Theory and Practice

When it comes to learning special or foreign languages, semantization is regarded as one of the most fundamental steps in mastering any language (Schmitt, 2008). Beheydt defines semantization as "a continuing process of getting acquainted with verbal forms in their polysemous diversity within varying contexts" (1987, p. 55). Over the years there have been various theories and frameworks, especially in the context of L2 which remain pertinent to the purpose of this study. As early as 1942, Cronbach identified two success factors for semantization: knowledge of word meaning, and accessibility to this knowledge. Dale in 1965 presented a progressive scale of semantization with four stages of semantization, ranging from the initial encounter with the word to mastering the word (Figure 2.1).

Four Stages of Semantization

(1) Never saw the word before.
(2) Heard it, but do not know what it means.
(3) Recognizes the word in context.

(4) Knows the word well.

Figure 2.1. Dale's Four Stages of Vocabulary Knowledge (Poole, 2011).

Paribakht and Wesche (1993) developed the Vocabulary Knowledge Scale with five levels that "range from total unfamiliarity through recognition of the written word and some idea of its meaning, to the ability to use the word in a sentence" (p. 15). The intent behind this model was to assist self-learners to assess and report meaningful gains in their vocabulary (Figure 2.2).

The Vocabulary Knowledge Assessment for Self-Learners

(1) I don't remember ever seeing this word.
(2) I have seen this word before but I don't know what it means.
(3) I have seen this word before and I think it means _____.
(4) I know this word. It means _____.
(5) I can use this word in a sentence. For example: _____.

Figure 2.2. The Vocabulary Knowledge Scale by Paribakht & Wesche (Poole, 2011).

Towards the end of the twentieth century, researchers started to entertain the idea of vocabulary learning as a continuum rather than discrete stages. A typical continuum would have little or no knowledge of a word at one end and expert knowledge at the other. One such progressive model called the Receptive/Productive Continuum was theorized by Waring in 2002 (Figure 2.3). In this model, the *receptive vocabulary* entails perceiving the proper sense of a word in one's mind but not necessarily expressing it; whereas the *productive vocabulary* is the measure of one's ability to express the correct sense of a word through speaking and writing.

Receptive Vocabulary   Productive Vocabulary

Figure 2.3. Waring's Receptive-Productive Vocabulary Continuum (Poole, 2011).

In general, receptive vocabulary is larger than the productive vocabulary in foreign language learners (Burger & Chong, 2011). Furthermore, learners with larger receptive vocabulary are likely to have a larger productive vocabulary (Webb, 2008). Poole (2011) reports that the one-dimensional vocabulary semantization continuums were later expanded to induct more dimensions. Nation (2001) posited that semantization is an interplay of knowing the grammatical forms of the words, their meanings, and usage. Daller, Milton, and Treffers-Daller (2007) extended Nation's work by proposing three components of semantization: breadth, depth, and fluency.

These frameworks and theories have greatly influenced the practice in the domain of L2 semantization. This influence is not surprising as Beheydt (1987) points out that semantization is anchored in applied linguistics and learning psychology and the two disciplines often come together to define effective semantization strategies and techniques. One such effective technique is *glossing*, which is the practice of embedding brief definitions of content words in texts. Schmidt (2008) notes that "there are several reasons why glossing can be useful: more difficult texts can be read, glossing provides accurate meanings for words that might not be guessed correctly, it has minimal interruption to reading – especially compared to dictionary use—and it draws attention to words that should aid the acquisition process" (p. 351).

Glossing can either be concordance-based or glossary-based. In the concordance-based glosses, learners are encouraged to infer meanings of new and unknown words from sample corpus texts. In contrast, the glossary-based glosses provide the meanings of the new and

unknown words. While studies in the L2 literature have generally confirmed that the efficacy of glossing in enhancing learners' incidental vocabulary (Azari, 2012), the conclusions of the studies on comparing gloss modalities have been mixed. Such inconclusive results could be explained by the fact that such comparative studies have been undertaken on the pretext of proving the superiority of one gloss modality over the other. Furthermore, the local contexts of such localized studies remain diverse and vary depending on factors such as L1 and L2 languages involved, geography, age, education, and other demographic features of the target population.

A few studies have also focused on ascertaining learners' attitudes when it comes to semantization. In one such study, nearly two-thirds of learners attributed the ease in comprehension to the availability of electronic glosses (Lenders, 2008). Lenders also reported that L2 learners use glosses not just to obtain new vocabulary but also to confirm their knowledge. Breyer (2009), however, noted that creating online glosses requires corpus processing skills that most language teachers lack.

The use of information technology in developing specialized software agents to automate tasks in humanities research such as corpus processing is at the core of *digital humanities* (DH). DH emerged in 2004 as a successor to *humanities computing* in academia to formally acknowledge the growing number of collaborative humanities projects using digital and multimedia technologies (Fitzpatrick, 2012). Since then, DH has served as an incubator for interdisciplinary collaboration. It brings together scholars and practitioners from humanities as well as computer and information sciences to work on core questions in each of these disciplines, such that "on the one hand semantic technologies are applied in novel ways in addressing

research questions of humanities and social sciences; on the other hand, these areas stimulate the development of novel methods in computer and information sciences." (Horrocks, Kagal, & Hotho, 2017).

Digital Humanities: Confluence of Computing and Humanities

In its simplest form, DH is defined as a field to "use digital tools to explore humanities questions" (Posner, 2018, p. 1). One of the areas that helped define digital humanities as a discipline is the analysis of literary and linguistic texts. In DH, however, the definition of "textual landscape" is expanding to include multimedia and digital sources. The term "digital text" is often used, in contrast to print text, to include digitized versions of print texts as well as born-digital materials. Jewitt and Kress (2003) coined the term "multimodal literacy" to advocate the need to incorporate various human modes (such as speech, gesture, vision, etc.) that are commonly used in communicating and representing ideas and knowledge. Walsh welcomes the idea of *multimodal literacy* to embrace "changes that have occurred with the range of digital media that are becoming embedded in people's lives" (2010, p. 212).

The emergence of DH as a discipline has also disrupted the traditional notion of publishing in humanities. This is because the significant role of data and computing in digital humanities has made it impossible for the print-era publication framework to support traditional processes (workflows, peer-review, marketing, etc.) to keep up with the pace of publication in DH (Muñoz, 2013). Muñoz further asserts that the *data curation* framework used in the scientific community should also be utilized by the humanities community; that is, "active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and

education…activities [which] enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time" (2013, p. 14).

Close versus Distant Readings of Texts

The notion of *distant reading* also surfaced in DH literature to differentiate data-driven and automated processing of digitized texts from *close reading* traditionally done by scholars as part of their research (Murdock, Allen, Börner, Light, McAlister, & Ravenscroft, 2017). Moretti (2013) argues that natural human capacity and speed for reading and processing texts are quickly becoming an impediment to the creation and discovery of new knowledge; scholars and practitioners need to incorporate distant reading in their scholarly endeavours to mitigate this problem. Seadle (2016) embraces the technology-mediated enhanced ability for researchers to undertake distant reading to overcome human limitations "with tools that to a large extent already exist and can become better as the academic community learns how to work with new techniques and to structure the content" (p.173). It can be argued that distant reading can also enable scholars to have more time to bring experiential and emotional interpretations to otherwise algorithmically "semanticized" data. This is the reason, as some scholars argue, that "digital humanities remains a fundamentally hermeneutic enterprise, and one in which distant readings and close readings must be tightly linked if anything is to make sense" (Murdock et al., 2017).

The Human-Machine Coexistence

A few scholars have sounded the alarm for academics to remain vigilant and critical about the impact of computing on the traditional discipline of humanities and also on the experience of being human. Flanders (2013) describes this as the "productive unease" of digital

scholarship that stems from the gulf between the *objective reality* and its *surrogate representations* by humans. Unsworth (2002) describes this dichotomy in DH as follows:

> Humanities computing, as a practice of knowledge representation, grapples with this realization that its representations are surrogates in a very self-conscious way, more self-conscious, I would say, than we generally are in the humanities when we "represent" the objects of our attention in essays, books, and lectures. (p. 1)

Izutsu (2002) expands on this idea by noting that our propensity as humans to attribute judgement through representative words to describe objects and experiences is part and parcel of the human experience. He asserts that languages have embedded worldviews that require careful semantic analysis to be made explicit. He illustrates this point with the example of the word "weeds" which conveys the value-laden meaning of *undesirable herb*. He argues that the connotation of being desirable or undesirable is a human construct and is not part of the *objective reality* that is being described. In nature, Izutsu posits, *weeds* remain equally valuable as *roses*.

Thus, with advances in computing, there is a risk of increasing this gulf between the objective reality and semantic representations of it by adding additional degrees of separation through the use of digital surrogates. In providing a critique of DH, Strongman (2008) pits humans against machines by asking: "Whilst [machines] may provide unexpected results, how insightful is it if the computer is responsible for the findings?" (p. 8). Flanders (2013) also supports the notion of humans guiding the humanities research rather than being driven by the fascination of what computing can do for humanities.

Given that DH is in its nascent stage as a discipline, these questions and criticisms should serve as warnings rather than warrants to cease human and machine collaboration in DH. In

responding to Flanders' idea of "productive unease," McCarty (2015) posits that DH is productive precisely because it makes some scholars uneasy as it forces them to find an inconvenient compromise between the way machines work and the way humans think. McCarty suggests that the trajectory of DH as a discipline will depend on how successful scholars in particular, and humans in general, are in addressing the "human-machine confrontation" that is at the core of today's digital age.

Summary

The literature review examined three primary strategies for preserving mixed languages including corpus processing, semantic analysis, and language learning and semantization, as well as digital humanities.. The best practices of corpus processing to develop lexicons were identified in the corpus linguistic literature. The common approaches of semantic analysis and semantization were examined from the perspective of language learning and digital humanities. The pragmatic literature review, as illustrated in the figure below (Figure 2.4), led to the notion of glossing in the L2 literature. The methodological approach taken to address the research questions underpinning this literature review is discussed in Chapter 3.
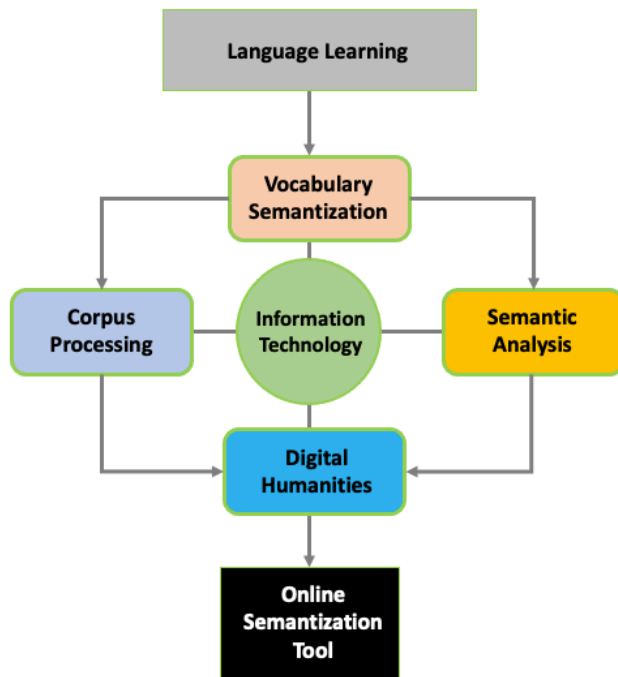
Figure 2.4. A Visual Summary of Literature Review

The multidisciplinary literature review undertaken for this study highlighted the contributions and evolution of information technology in the domain of language preservation, albeit within the protection and comfort of their respective disciplinary boundaries and traditions. While this plurality of disciplinary traditions and perspectives is healthy in approaching societal issues, some problems require disciplinary researchers and domain experts to collaborate across disciplinary boundaries to find interdisciplinary solutions. Klein & Newell (1997) define interdisciplinary solutioning as "a process of answering a question, solving a problem, or addressing a topic that is too broad or complex to be dealt with adequately by a single discipline or profession" (pp. 393-394). This study argues that safeguarding mixed and endangered requires interdisciplinary perspective using DH, which also necessitates a commitment to the use of information technology in designing solutions.

Situating the Semantization of Ginans

The learning and understanding of a language play a crucial role in safeguarding traditional knowledge that are oral and mixed such as the one used in the ginans. This research is based on the premise that ginans can continue to live in the diaspora if the successive generations are motivated and able to learn and understand its language. This study frames the problem of learning the language of ginans as equivalent to that of learning a foreign language by online learners whose native language is English.

As discussed in the literature review, the process through which humans learn words of a language and their meanings is called semantization. The emerging evidence in L2 literature suggests glossing to be an effective technique to facilitate vocabulary semantization for learners. However, a common measure to evaluate the efficacy of glossing is the extent of learners' productive vocabulary (i.e., the words that are used by learners) as opposed to their receptive vocabulary (i.e., the words that can be understood but not necessarily used by learners). This metric, however, is of little use in the context of the language of the ginans.

The historical practice of composing ginans came to an end in the mid-nineteenth century and no new ginans have been composed since then (Asani, 2011). The learning goal of this study, therefore, differs from other L2 semantization studies due to its focus on developing learners' receptive vocabulary rather than their productive vocabulary. This deviation is justified as the community does not expect its youth to learn the language to compose new ginans. Instead, the current community efforts are focused on safeguarding the existing ginans through transmission and teaching as opposed to creating of new ones.

CHAPTER 3: RESEARCH METHODOLOGY

This chapter outlines both the research methodology and the underlying prototyping method used in this study. It provides a brief introduction and rationale for the use of design-based research (DBR) methodology followed by a detailed discussion on the use of the educational research and development (ER&D) method for prototyping the online ginan semantization tool.
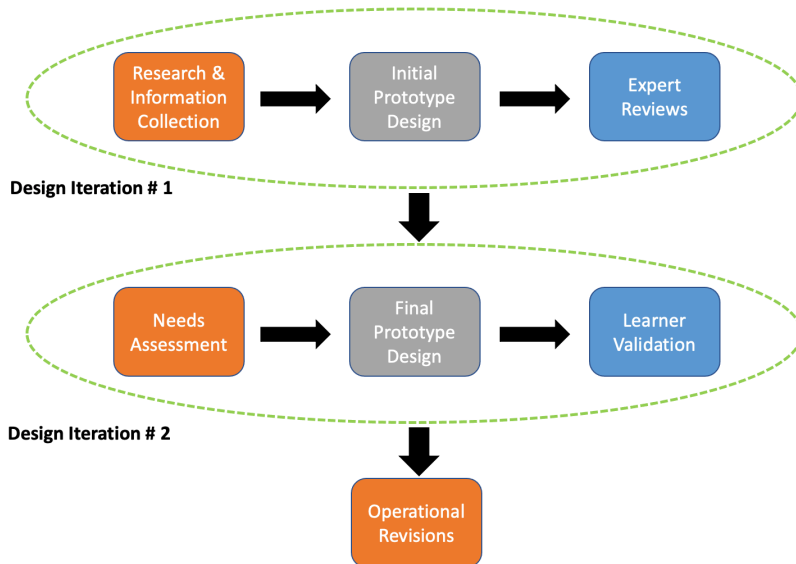
Design-Based Research Methodology

This study adopted design-based research (DBR) methodology to prototype an online ginan semantization tool. Semantization in the context of this study refers to the ability to learn and understand the vocabulary of ginans. The DBR methodology is particularly well-suited for practice-based disciplines, such as education and librarianship, to derive evidence and principles for developing theories and designing interventions to improve practice. Whether utilized to improve practice or inform theory, the research typically involves "an iterative process of successive approximation or evolutionary prototyping of the ideal intervention" (Van den Akker 1999, p. 8).

DBR methodology works particularly well for designing educational prototypes and products "through iterative analysis, design, development, and implementation based on collaboration among researchers and practitioners in real-world settings" (Wang & Hannafin, 2005, pp. 6-7). When developing prototypes using DBR, the initial iteration begins with a theoretical design grounded in the literature, which is refined with subsequent iterations based on input from educators and learners in the field (Holmes, 2013).

Educational Research and Development Method

A popular process for prototyping educational resources and interventions is Educational Research and Development (ER&D). The ER&D method found wide acceptance as a prototyping method to "endure a cycle of field-testing, evaluation, and refinement for producing functional educational products and programs" (Pownell, 2002, p. 45). Consistent with the tenets of the DBR methodology, ER&D is also a continuous and iterative design process with seven distinct steps, which are: 1) literature and information collection 2) needs assessment, 3) initial prototype design, 4) expert review and consultation, 5) final prototype design, 6) learner validation, and 7) operational revisions (Borg & Gall, 2003). The design-based prototyping of the online ginan semantization tool was spread across two design iterations (see Figure 3.1).

Figure 3.1. The Design-Based Research Methodology Used in this Study

The first iteration focused on designing a proof-of-concept of the tool informed by theoretical best practices in semantization literature as well as practical considerations of building a corpus-based lexicon as discussed in Chapter 2. In the second iteration, the initial

design was evolved into a functional prototype through successive refinements guided by the feedback received from experts and potential users.

Step 1: Literature and Information Collection

The goal of this study was to design an online prototype for ginan semantization using the available ginan materials and resources (or corpus). As has been noted previously (Chapter 2), glossing is an effective technique for enhancing semantization for L2 learners. There are two common modalities of textual glossing that have been used in L2 research studies, which are: dictionary-based (meaning-given) and corpus-based (meaning-inferred). Rather than taking a comparative research stance, this study is based on the premise that with the use of information technology, it is possible to provide both gloss modalities to learners simultaneously. To this end, in the first step of the ER&D prototyping process, a technology-based semantic analysis was undertaken to create an online lexicon to facilitate glossing of the sample ginan corpus.

Sample Ginan Corpus and Lexicon. As discussed in Chapter 2, the type of corpus to be used depends on the language and the purpose of the study. Given that the mixed language of ginans is primarily Indo-Aryan with loanwords from Perso-Arabic languages, a parallel corpus with Romanized texts and English translations was chosen for this study. As noted previously, the production of ginan literature remains dispersed among community individuals and academic scholars. A further review of the available publications in these two categories revealed variant and incompatible use of romanization conventions across the materials. Consequently, materials from different authors could not be mixed and matched for this study. As a result, textual nd lexical ginan materials produced by Kamaluddin Ali Muhammad remained the only viable option for creating a parallel corpus needed for this study. It is important to recall that the scope

44

of this study requires the use of the existing ginan materials available to the community for the prototype to be of practical use. Thus, the creation of the sample parallel corpus for this study was based on pragmatic considerations such as the availability and combability of the representative ginan resources. The names of the resources inducted in the sample ginan corpus for the study along with the list of 10 ginans are provided in Appendix A.

      Developing an Online Ginan Lexicon. The following semantic analysis process was applied to the sample ginan corpus for developing an online ginan lexicon.

1. *Identifying candidate terms:* A candidate term is a term that has been identified, either manually or through means of automated process, as one with the potential to qualify for inclusion in the final lexicon. It is a basic meaning-bearing lexical unit of a language and may consist of one or more words. For this study, all the terms in the concordance of the 10 sample ginans were designated as candidate terms.

2. *Selecting content terms:* The content terms are words (nouns, verbs, adjectives, and adverbs) that are chosen to be included in the lexicon from the pool of candidate terms either manually or via an automated algorithm. The candidate terms for this study were identified using an automated algorithm. This algorithm matched the concordance terms with the terms identified in the glossary of the 10 sample ginans chosen for this study. A successful match between the concordance and glossary terms was used as the criterion for the selection of content terms.

3. *Identifying word senses:* Since words are generally polysemous (having many meanings or senses), the fundamental principle of semantization is the ability for learners to identify and understand various senses of the terms. For this study, *word*

sense disambiguation (WSD), which is the process of "determining which sense of a word is being used in a particular context" (Jurafsky & Martin, 2019, p. 354), was utilized. It entailed matching and retrieving word senses of the English synonyms of the candidate terms from the Online WordNet database.[10]

4. *Developing an online semantic profile:* An online semantic profile describing various linguistic characteristics of each content term (e.g., parts of speech, word senses, synonyms, etc.) was developed in this study. These semantic profiles were also integrated with the underlying corpus sources as depicted in the figure below (Figure 3.2).
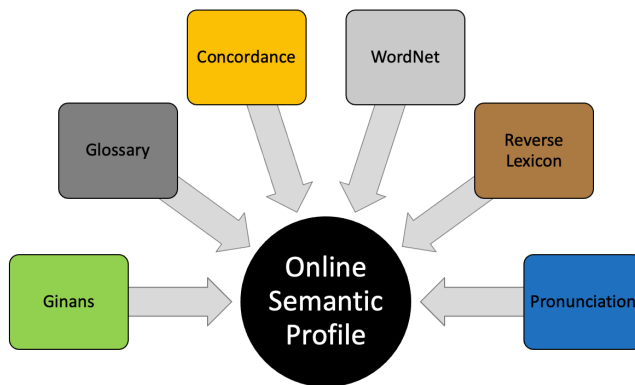

Figure 3.2. Integrated Online Semantic Profile of Content Terms

As has been mentioned before, the overall purpose of undertaking a semantic analysis of the ginan corpus was to provide learners with both concordance-based and glossary-based glosses to facilitate semantization. The results of the semantic analysis are discussed in Chapter 4.

---

[10] Online WordNet® is a large lexical database of English developed and hosted at Princeton University. It can be accessed at https://wordnetweb.princeton.edu/perl/webwn

Step 2: Needs Assessment

The purpose of this step was to gather the needs of learners who are interested in understanding the vocabulary and meanings of the ginans. While the semantization tool was developed in close collaboration with ginan experts, it was also important to solicit and analyze the needs of the learners for this study. The target population for this study is defined as the members of the Ismaili community between 18 to 44 years of age living in Euro-American countries whose primary language is English. The needs of the target population are distinct from those of their elders who migrated from their native land to Euro-American countries. An online survey was developed to solicit the needs of the target population.

Pilot Survey. The initial draft of the survey covered demographics, attitudes towards the ginans, and the use of information technology and tools to access ginan materials.[11] The survey questionnaire was reviewed by the research supervisor as well as a survey specialist at the University of Saskatchewan's Social Sciences Research Laboratories (SSRL) for consistency and quality. The two reviews resulted in refinements to the draft questionnaire. For instance, the initial questionnaire was modified to include three new questions on gender, occupation, and language fluency. In addition, the survey was made more readable by using consistent spelling and formatting throughout the survey. The revised survey was then administered to a small sample of the population as a pilot. The local members of the Ismaili community in Saskatoon were invited via email to complete the online pilot survey. A total of five 5 members completed the pilot survey as well as the post-survey feedback. This feedback was gathered utilizing a questionnaire on the readability and language of the main survey.

---

[11] The online survey questionnaire was developed and administered using the SurveyMonkey platform which is supported by USask and available to all students without any additional cost.

Final Survey. The final online survey was promoted to the community through various websites and social media platforms, including the Ginan Central portal. This portal is hosted by the University of Saskatchewan Library and attracts people interested in the ginans from across the globe. At the beginning of the online survey, participants were presented with a brief introduction covering the purpose of the study. After the introduction, the participants were presented with an online consent form to read and agree to. Following the consent, the participants were presented with the needs assessment questionnaire. None of the 22 questions of the survey was made mandatory, allowing participants to skip over any of the questions. The consent and the main survey questionnaire are available in Appendix B.

Recruitment of Survey Participants. Although the survey was available publicly to all adults aged 18 years old and above, the study utilized convenience sampling of participants from the Nizari Ismaili community. The members of the Ismaili community were informed of the online survey using a variety of communications channels popular in the community, including blogs, websites, as well as social media. For instance, the survey was advertised on Ismailimail, which is a popular blog in the community.[12] Several websites dedicated to the ginans were also contacted to inform their respective patrons about the survey, including Ginan Central at the University of Saskatchewan, Ismaili.Net Heritage, JollyGul, and GinanGuru. Professional and personal contacts of the researcher were also informed about the online survey via e-mail and encouraged to spread the word to their contacts in turn. The results and analysis of the evaluation will be discussed in Chapter 4.

---

[12] The survey invitation on Ismailimail can be accessed at https://ismailimail.blog/2020/07/22/ginan-survey-2020-by-ginan-central-at-the-university-of-saskatchewan/

Step 3: Initial Prototype Design

The focus of this step in the overall research process was to design an initial design or a proof-of-concept for the online ginan semantization tool. The design at this stage of the research was essentially a conjecture informed by the review of the existing semantization literature and the resulting semantic analysis of the sample corpus used for this study (Step 1).

Design Process. The initial design of the online ginan tool was generated using *wireframe*. A wireframe visually brings together the content, interface, and navigational features of a system or tool. While wireframes can be drawn on a piece of paper or whiteboard, a mock webpage was used for this study. The four resources utilized and integrated for the initial design included ginan texts, translations, tunes, and semantic profiles or lexicon of the content words (Figure 3.3). The goal of the initial design was to integrate and visually present all four underlying resources to facilitate semantization based on the best practices derived from the literature. As part of the ER&D prototyping process, the initial design was then presented to community experts for review and consultations.
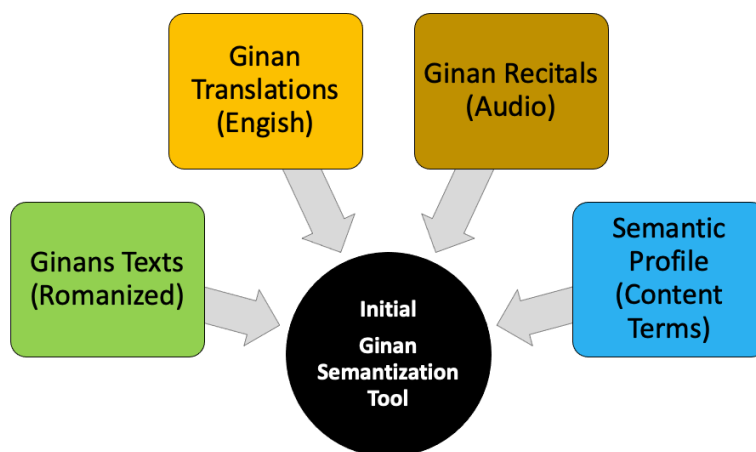

Figure 3.3. Elements of the Initial Design of the Ginan Semantization Tool

Step 4: Expert Reviews

The iterative ER&D process also calls for the initial design to be reviewed by experts to solicit initial feedback before proceeding to the next iteration. To this end, the initial design underwent expert reviews to ensure that the perspectives and practices of the community elders and educators were captured and incorporated before proceeding to the final design iteration.

Experts Selection and Recruitment. The purposive selection of experts for the review was based upon the following criteria:

a) A member of the Ismaili community, who is

b) currently involved in transmission and/or teaching of ginans,

c) knowledgeable about the issues and challenges faced in the teaching of ginans, and

d) agrees to participate in the tool evaluation by signing informed consent forms.

The experts were approached via email and asked to provide verbal consent at the beginning of their sessions. The text of the invitation email and expert consent form is provided in Appendix C. The three reviewers were selected using purposive sampling method. The names of the three experts have been encoded to ensure privacy and confidentiality. The demographic information of the selected expert reviewers is presented in the table below (Table 3.1).

Table 3.1

Community Experts Recruited for Reviews

	Location	Age Range (years)	Gender
Reviewer A	Vancouver, BC	60 - 70	Male
Reviewer B	Calgary, AB	30 - 40	Male
Reviewer C	Edmonton, AB	50 - 60	Female

The demographics of the reviewers are representative of the wider community demographics where the majority of the ginan experts and educators remain accessible to communities in bigger cities. In smaller cities like Saskatoon and Regina, community members do not have in-person access to ginan educators. This situation further justifies designing a solution that transcends geographical barriers using information technology.

Review Process. A virtual and one-to-one session was held with each of the three experts using online video conferencing tool.[13] On average, these review sessions lasted between forty-five and sixty minutes and were divided into the following parts.

Research overview and consent. At the beginning of the session, each expert was given an overview of the research and briefed on the purpose of the study. After the briefing, each reviewer was allowed to ask questions until they felt comfortable and confident about their participation. The participants were then asked to provide consent and were reminded about their right to request for the recording to be turned off at any time.

Self-introduction. Following the consent, the participants were asked to briefly introduce themselves and share their general practice of teaching ginans and the ginan resources they use.

Review questions. Each participant was also given a demonstration of the tool's interface, highlighting its layout, features, and functionalities. The demonstration was followed by an interview of the expert geared toward a review of the initial design. The list of guiding questions used during the review is presented in Appendix D. The completion of expert reviews marked the end of the initial design iteration. The feedback from the review, together with the analysis of

---

[13] The online conferencing tool used for this study was Webex, which is supported by USask and available to all students without any additional cost.

learners' needs assessment (Step 2), served as input for the design of the prototype. The results and analysis of the evaluation will be discussed in Chapter 4.

Step 5: Final Prototype Design

While not a full-fledged product, the final design of the prototype incorporated functional and technical design considerations necessary to support and sustain an operational tool. The functional design considerations include aspects of placement and prominence of content, user interface, and navigational aspects. The technical design considerations encompass the necessary technological infrastructure, including database design, data upload, metadata coding, webserver, and software.

Design Process. The feedback received from the reviewers was used as the basis for the final design of the prototype. This approach is consistent with the ER&D process as it allows iterative and continuous improvements of an educational product based on formative evaluation (Borg and Gall, 2003). The transformation of the initial design marked the beginning of the second design iteration. In addition to the formative evaluation by experts, the needs of the learners obtained through the online survey were also incorporated in the final prototype design process. The information from the online survey and expert evaluations were analyzed to identify features for the prototype. These features were then categorized into three categories: 1) content, 2) interface, and 3) navigation. The features list was ranked based on support from experts and learners and the selection of final features was made based on feasibility criteria.

Feasibility criteria. The feasibility criteria helped to find a practical balance among three essential parameters: 1) time required to implement a feature, 2) resources available to

implement a feature, and 3) technical complexity of a feature. The revisions made to transform the initial design during the second iteration are discussed in Chapter 4.

Step 6: Learner Validation

Learner validation of the prototype determines whether the prototype has met the needs of the learners as an effective online ginan learning tool. During this step of the design iteration, the intended target users of the tool were invited to use the prototype and evaluate it by participating in the validation survey.

Selection criteria. The ER&D process requires the prototype to be validated by its intended users. Using the convenience sampling method, an email invitation was sent to all the members of the target population who provided their contact information during the needs assessment survey. This sampling approach ensured that the prototype incorporated perspectives of the target population identified for this study. The potential participants were given the necessary background and access information about the survey via email.

Evaluation Process. The learner validation was a two-step process. In the first step, the participants were given a link to access the prototype and asked to familiarize themselves with the tool before the assessment. To aid participants in their exploration and evaluation of the prototype, the following strategies were also suggested to the participants:

1. Explore the ginans that are most interesting, and perhaps even unfamiliar, to you from the available list of sample ginans.

2. Click on the "Back" button to return to the ginan list.

3. Explore various options in the toolbar by clicking each of the buttons.

4. Listen to the recital of your chosen ginan by clicking the play button of the audio player.

5. Access the pronunciation of special letters (such as ā, ī, ū, etc.) by clicking on the Pronunciation Guide link.

6. Review the side-by-side and verse-by-verse text and translation of the ginans.

7. Read the English text of the ginans and hover over and click on the linked words to access more information.

8. Explore the semantic profiles of various linked ginanic terms.

9. Learn to pronounce difficult words by clicking on the speaker icon in the semantic profile of the words.

10. Read the English meanings and synonyms of important terms in the ginan presented in the glossary.

In the second step of the learner validation process, the participants were asked to fill out a short online survey to provide their feedback on the prototype. The survey questionnaire consisted of seven questions to solicit feedback on the content, interface, and navigation features of the prototype. One open-ended question inviting other comments on the prototype was also included. The validation survey questions are included in Appendix E.

Validation Survey. The validation survey was also pilot tested before its final deployment. Based on the feedback of the pilot, several changes were made to the prototype and the questionnaire. For instance, all references to the term "semantization" were replaced with the term "learning." This change was warranted as neither of the two pilot participants was familiar with the term.

The final survey was sent to the participants on November 28, 2020. Each participant was provided with a unique link to the online evaluation survey.[14] The validation data from the participants was analyzed quantitatively to ascertain the overall efficacy and acceptance of the prototype. The open-ended comments were also analyzed to guide any necessary revisions to the prototype, which will be discussed in Chapter 4.

Step 7: Operational Revisions

The feedback received from the prototype validation survey (Step 6) was assessed to inform the final revisions to the prototype. Given that a high percentage of respondents preferred to use smartphones to access the tool, there was a suggestion to convert the tool into a mobile app from its existing web version. While creating a mobile app was not feasible for the scope of this research, the survey interface was updated to be mobile-friendly.

---

[14] The validation survey was developed and administered using the SurveyMonkey platform which is supported by USask and available to all students without any additional cost.

CHAPTER 4: ANALYSES AND RESULTS

This chapter discusses the results of the various analyses undertaken as part of prototyping the online ginan semantization tool. It begins with the discussion of the semantic analysis undertaken during the research and information collections process. The initial design of the prototype is discussed next, followed by the analysis of expert reviews of the initial design. A significant portion of this chapter is dedicated to the analysis and discussion of the results of the learners' needs assessment survey. Next, the design features of the final design of the prototype and its evaluation are discussed. The chapter concludes with a discussion on some of the considerations for using the online ginan tool in the community.

Semantic Analysis

As discussed in Chapter 2, online glossing is a key enabler of L2 semantization, which in turn depends on the availability of electronic lexicon. To facilitate online glossing of romanized ginan texts, the printed ginan concordance and glossary materials chosen for this study were remediated into electronic database. The semantic analysis of the sample ginan corpus comprised four steps (see Figure 4.1), which are discussed next.
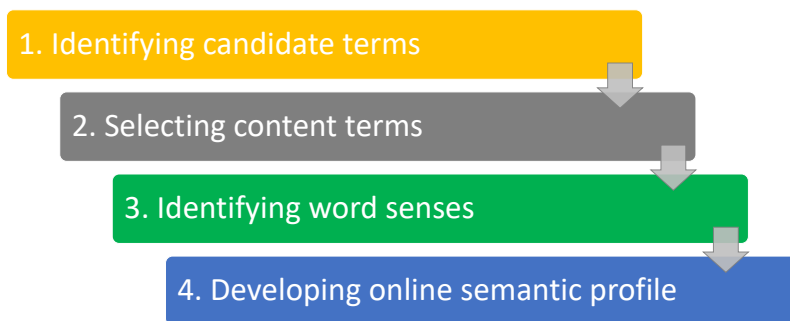
Figure 4.1. The Semantic Analysis Process Used in the Study.

1. Identifying candidate terms. A candidate term is a term that may potentially qualify for inclusion in a lexicon. Using the unstructured semantic analysis process, the metaphorical "bag of words" was filled with nearly 3,000 unique terms that were extracted from the concordance volume chosen for this study. After filtering the corpus bag-of-words for the 10 ginans chosen for this study (Appendix A), there were 347 unique candidate terms, which were collectively used 479 times in the chosen sample of 10 ginan texts.[15]

2. Selecting content terms. Content terms are words (nouns, verbs, adjectives, and adverbs) that have a clearly defined meaning. The glossary resource in the sample corpus was used for identifying content terms by matching them with the available candidate terms. The text-based matching of the terms was automated using MySQL database, and an open-source relational database management system, using structured query language or SQL. The automated process resulted in the selection of 119 content terms, which amounts to 34% of the candidate terms.

The automated matching algorithm was further enhanced by overriding the default match settings to ignore diacritics on letters such as ā, ī, ṭ, and others. This override allowed letters with diacritics to be considered equivalent to their counterparts. For instance, this override allowed the word *karma* to be same as *kārmā* or *karmā* during the matching process. This flexibility in character matching is particularly important for the ginans due to the extensive use of variant romanization conventions. With this enhancement in the automated matching algorithm, the match rate jumped to 41% or 143 candidate terms from 119 or 34%. The table below (Table 4.1) provides a statistical summary of the automated matching algorithm used for this study.[16]

---

[15] The list is accessible at https://ginans.usask.ca/semantics/analysis/summary_conc.php.
[16] The list is also accessible at https://ginans.usask.ca/semantics/analysis/.

Table 4.1

Statistical Summary of the Automated Matching Algorithm[17]

	Pre-Override		Post-Override	
	Number	Percentage	Number	Percentage
Content (Matched) terms identified	119	34%	143	41%
Non-content (Unmatched) terms	228	66%	204	59%
Mismatched (False-positive) terms	0	-	7	2%
Total Candidate (Concordance) terms	347	100%	354	102%

An unwelcome consequence of this enhancement was the possibility of false-positive matches. A false-positive occurs when homonyms, i.e., two semantically different words, are matched as one and the same based on their romanized spellings. An example of false-positive would be the positive matching of the word *bhav* (meaning life) with *bhāv* (meaning nature). As noted in the table above, a total of one hundred and fifty content words were identified, including seven mismatches. With the false-positive rate of 2%, the benefit of overriding the automation default settings outweighed the risk of false-positive matches. The false-positive matches were manually corrected as part of the word sense identification step.

3. Identifying word senses. The next semantic analysis process was to correctly identify the word senses of the content terms. This task was achieved through the word sense disambiguation (WSA) algorithm. This algorithm entailed matching the English synonyms of the content words with WordNet, an online lexical database for English developed at Princeton

---

[17] Further details underlying the statistical summary are available at
https://ginans.usask.ca/semantics/analysis/summary_gloss2con.php

University. For instance, in the case of the word *bhāv*, its English synonym "nature" was used to

match and retrieve relevant lexical information from WordNet as illustrated in the figure below

(Figure 4.2). WordNet organizes words with the same word sense as synonym sets or *synsets*,

Each synset in WordNet is assigned a part of speech and a gloss, which provides a description

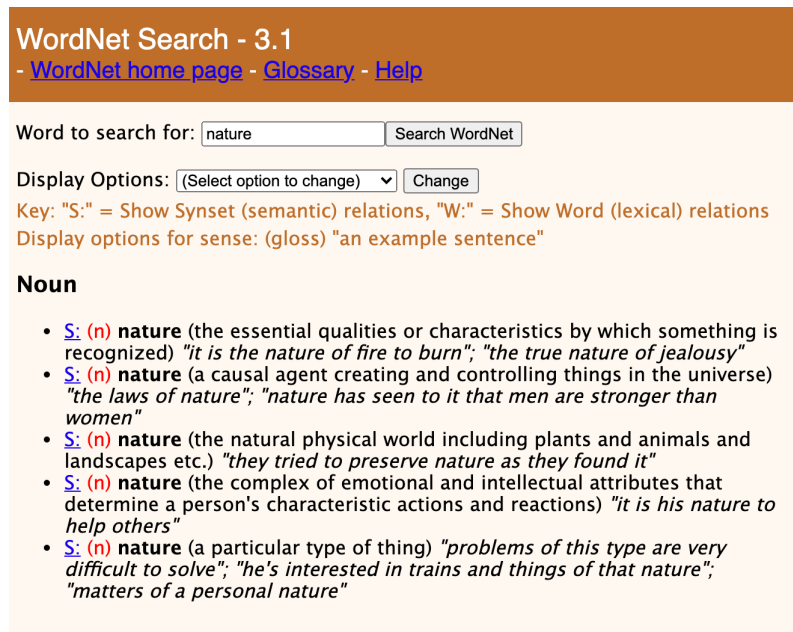and example sentences to convey the correct meaning of the word.

Figure 4.2. An Illustration of Synset from WordNet Online.

The ginan glossary used for this study provides a comma-separated list of English

synonyms for each term. These synonym lists were *tokenized* for matching and retrieving lexical

information from WordNet. Tokenization is a process of splitting meanings, often separated by a

delimiter such as a comma, into discrete words. Each token was then used to match against the

words in WordNet to retrieve the corresponding synset and gloss. In this study, only the initial

token of the content words was used for WSD to avoid duplication and exponential usage of

computing resources. The WSD algorithm resulted in matching 73% of the content terms.[18]

Upon manual verification, the total number of correct matches was reduced to 26%. This sizable

discrepancy confirms the polysemous nature of the ginan vocabulary and the need for developing

a richer ginan lexicons to improve automated word sense disambiguation.

A potential improvement that can be made is to categorize candidate terms based on parts

of speech (e.g., nouns, verbs, adjectives, adverbs, etc.). Since the synsets in WordNet are

organized by parts of speech, utilizing this information in the automated matching process could

improve the results. This change would, however, also require data identifying the parts of the

ginan vocabulary, which is not currently available for the ginans. Based on these constraints, no

further enhancements were made to the algorithm or the process.

Notwithstanding the lack of success in automating the process, WSD remains an

important semantic process for identifying the lexical field of the ginan vocabulary. The WSD

process for the current sample revealed that around 5% (n = 7) of the content terms were

semantically related based on word senses (see Table 4.2). This is a promising development as it

opens new avenues of research in distributional semantics for the ginan vocabulary.

Table 4.2

Examples of Semantically Related Content Terms based on Word Sense

Word Sense	Synset	Content Terms
god.n.01	God, Supreme Being	kirtār, sāmī, sirjaṇhār
heart.n.01	heart, bosom	haiḍe, man
together.r.01	together	sang, sāth

---

[18] For details on automated WSD, see https://ginans.usask.ca/semantics/analysis/details_gloss2wsa.php.

4. Developing an online semantic profile (lexicon). The semantic profile or lexicon created for this study enumerates and describes pertinent attributes of the content terms. A semantic profile template was designed to capture these attributes in a consistent fashion as illustrated in the table below (Table 4.3). Each content term is assigned a unique numeric identifier to avoid duplication of the same content with variant romanization and grammatical forms. The preferred romanization form of a term is recorded under *content term* in the template, while the variant romanization forms, if any, are captured in the *romanization form(s)* attribute of the template. The sound-based transcription recorded in the *phonetic notation* considers the number of syllabi and the diacritics used in the content words. It was important to use the Latin script in assigning phonetic notations to allow for familiarity and accessibility to English-speaking learners.

Table 4.3

Semantic Profile Template for Candidate Terms

Attribute	Description
Content term ID	A unique numeric identifier assigned to a term.
Content term	Content word or phrase in romanized form.
Variant form(s)*	Variant romanized transcriptions of the term, if any.
Phonetic notation*	A pronunciation guide of the term using the Latin script.
Part of speech	The part of speech of the term e.g., noun, verb, etc.
Gloss	A short sentence describing the sense of the content term.
English synonyms(s)*	Equivalent words in English for a given sense of the term.
Equivalent term(s) *	Other ginan terms that share the same meaning and sense.

Attribute	Description
Usage example*	Excerpts of ginan texts that use the content term
Related terms*	List of semantically related words.
Related ginan(s)*	List of ginans that use the content term.
Source(s)*	List of ginan materials referenced.

Note: * These attributes may be repeated more than once for a given content term.

The table below (Table 4.4) provides the phonetic notations for the common diacritics used in the romanized ginan corpus. Pronunciation audio files for over 80% (n = 120) of the content words were included in the semantic profile to aid in semantization.

Table 4.4

Phonetic Equivalents of Diacritics Used in Romanization of Ginan Corpus [19]

Diacritic	Phonetic Equivalent
ā	/aa/ to emphasize 'a' sound as in *car, jar,* etc.
ḍ	/dd/ to emphasize the 'd' sound as in door, or /dr/ to capture the combination of 'd' with the retroflexive 'r' sound as in *driver*.
ī	/ee/ to emphasize 'i' sound as in *teeth, cheer,* etc.
ḷ	/lr/ to capture the combination of 'l' with the retroflexive 'r' sound.
ṅ	/(n)/ to highlight the nasal sound in 'n' as in *sung, tongue,* etc.
ṇ	/nr/ to capture the combination of 'n' with the retroflexive 'r' sound.
ṭ	/tt/ to emphasize the 't' sound as in *tattoo, tap,* etc.
ū	/u/ or /oo/ depending on the usage as in *university* and *tool.*

[19] An online version of the table with corresponding digital sound for the diacritics is accessible at
https://ginans.usask.ca/semantics/prototype/pronunciation.php.

Although no such case was encountered in the sample corpus, the semantic profile template is designed to accommodate the possibility of having multiple senses for a given content word. Thus, the corresponding four attributes in the template (*part of speech, gloss, English equivalents*, and *equivalent terms)* are repeated for each sense of the content term encountered.

As mentioned before, the content terms which share the same gloss are listed as *equivalent terms*. The sample ginans that share a content term are identified as *related ginans*. The sources used to populate the attributes of a given content term are also noted for reference. The sources included in the sample for this study are listed in Appendix A.

Having a semantic profile template made it easier to store the data electronically to facilitate online semantization. The online lexicon data was stored using several tables in the MySQL and was made accessible on the Web using HTML markups and PHP scripting.[20]

Initial Prototype Design

As mentioned in Chapter 3, balancing the best practices of language learning from the literature as well as community elders was an important consideration in the initial design of the prototype. The design wireframe gathered the necessary resources in one place for a given ginan. These resources included romanized ginan texts, English translation, audio recitations, as well as glossary as illustrated in the figure below (Figure 4.3). The wireframe of the design went through several refinements as the desired functionalities and user-experience decisions were made. For instance, rather than listing all the recitals of the ginans, a specific recital conducive to learning was chosen for simplicity. A single link to other recitals was then presented for learners

---

[20] The online lexicon and semantic profiles can be accessed at
https://ginans.usask.ca/semantics/analysis/content_terms_index.php.

interested in listening to other renditions of the ginan. The initial design was used as a basis to solicit formative feedback from ginan reviewers who have experience in teaching ginans in the community.[21]
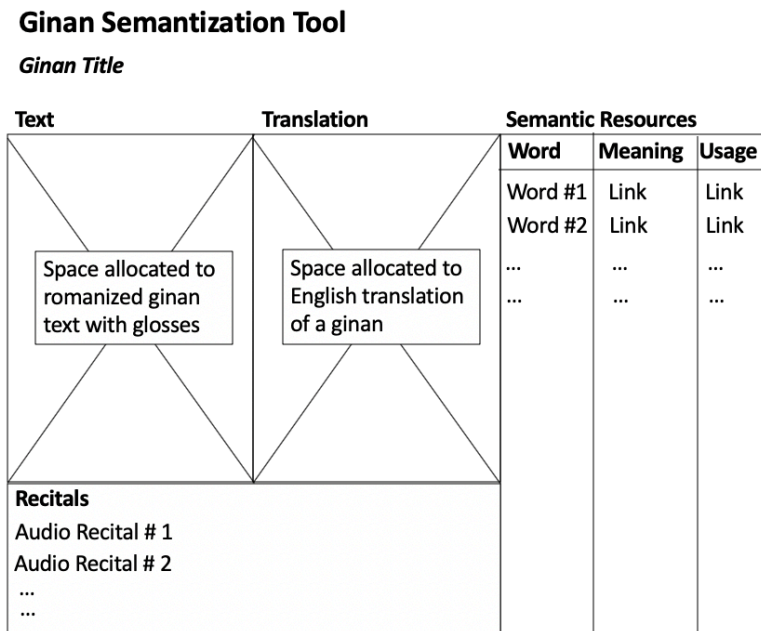
Ginan Semantization Tool

Ginan Title

Text	Translation	Semantic Resources		
		Word	**Meaning**	**Usage**
Space allocated to romanized ginan text with glosses	Space allocated to English translation of a ginan	Word #1	Link	Link
		Word #2	Link	Link
		…	…	…
		…	…	…
Recitals Audio Recital # 1 Audio Recital # 2 … …				

Figure 4.3. The Schematic of the Initial Design of the Prototype

Expert Reviews Analysis

In a typical in-person ginan learning session, the instructor provides a paper copy of the ginan texts and translations to their students. The instructor gathers the learning materials from various printed sources to share with learners. The initial design of the online semantization tool emulated the printed resources used in the in-person ginan sessions. It came as no surprise that the collocation of ginan texts and translation in the initial design was deeply appreciated by the three experts independently and unanimously. This design strategy also brought a degree of

---

[21] The initial design used in this study can also be accessed online at https://ginans.usask.ca/semantics/poc.

familiarity and comfort to them in the face of a deep paradigm shift from in-person sessions to online ginan learning espoused by the prototype.

The formative feedback solicited from the experts fell into three categories: content, interface, and navigation. The content category was used to gather feedback on the nature and types of resources for the semantization tool. The feedback and feature requests on the user interface of the semantization tool were filed under the interface category. Finally, the navigation category was used to group experts' feedback and impressions on the ability of the semantization tool to integrate and add additional links and buttons to optimize navigation to and from other resources and information.

Content. Since the prototype design considered the type of content currently being used in in-person instructional sessions, all three experts agreed on the use of ginan text, translation, and audio recordings as the core resources needed for learning and understanding ginan. Furthermore, the inclusion of the online glossary and the integration of the online semantic profile as part of the tool was unanimously regarded as useful content. Two of the three experts suggested incorporating composer information as well as including a gist or the summary of the ginans as additional content for the tool. On the topic of the online glossary, all three experts agreed with the observation that the creation and integration of such a tool is a unique benefit to learning and understanding of ginans online.

When asked about the inclusion of ginan videos as a possible content type for the tool, all three experts strongly opposed the idea. The rationale given for discouraging the use of ginan videos in learning ginan is the lack of control that instructors have in navigating the content easily and accurately. One of the experts explained that ginan videos could be effective in

community gatherings but not for teaching ginans. In an instructional setting, the instructors must have the autonomy and ability to adjust the pace of instructions based on the needs and skills of the class, which is generally hard to achieve with pre-recorded videos. An additional disadvantage of using videos is that ginan texts and translations cannot be viewed in their entirety, which may limit learners from reading ahead or referring to familiar content. The strong reservation on the use of videos in in-person sessions by the reviewers is understandable given these sessions are led by an instructor. However, this reservation may not be as relevant when ginan learning occurs in an online and self-learning environment.

Interface. There were several suggestions made by instructors to improve the overall interface of the tool. Two of the three experts independently suggested integrating ginan categories into the interface, so the learners can learn and lookup other ginans in each category. Another common suggestion was to allow electronic marking of the ginan texts to aid in memorizing the tune and cadence of reciting ginans in prescribed tunes. Historically, the ginans were transmitted in prescribed tunes that match the message and mood of the individual ginans. In an in-person instructional setting, it is a common practice for instructors and learners to mark their printouts of ginan texts with symbols like bars and dots to capture the necessary grouping and inflections of words when learning to recite ginans in prescribed tunes. One of the experts shared their markings of a ginan text for reciting ginans (Figure 4.4). These markings are not standardized and vary from individual to individual to serve as aide-mémoire.

Eji Sahentar deepme(n) Shah jaaher betha
Saanj sibhu nit jampo mora bhai........Arti kije naklank tañi ji

Satgur sohodeve raa bataai
Baar gur Pir Sadarhine dish bataai........Arti kije naklank tañi ji

Eji Ajampiya jaamp munivare haaniya
Lakh chorasi nivna fera taaliya.....Arti kije naklank tañi ji
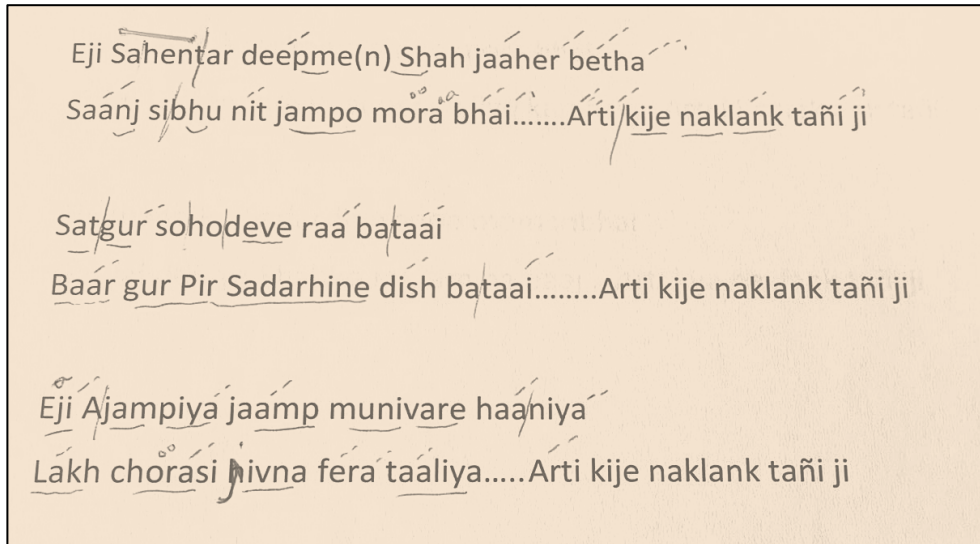
Figure 4.4. An Illustration of Personal Annotations for Learning Ginan Tunes

Additional individual suggestions included linking the tool interface with other ginan resources available on online portals such as Ginan Central. One expert suggested highlighting ginans that are rare or are on the verge of being forgotten. Doing so, according to the expert, will not only help identify at-risk ginans but may incentivize learners who often strive to master unique ginans to increase their chances of being called upon to lead congregational recitals. A technical suggestion made with regards to the interface was to enhance the audio player in the tool for better precision to make it easier for the users to play and replay parts of recitations using the slider while learning difficult verses. By far the most creative suggestion by an expert was to incorporate the "Ask-an-Instructor" feature to make it easier for online learners to connect with ginan experts and instructors if they need help.

Navigation. Several useful navigational features were also suggested during the individual review and consultation sessions with the experts. One suggestion was to add a special print button that enables users to print the text and translation of the ginan text rather than the

entire screen. One expert also suggested adding a special download button that lets users download ginan text and translation as a pdf document.

Overall, the experts were very supportive of the general concept of the initial prototype and offered favourable comments and suggestions for improving the content, interface, and navigation to transform it into a prototype.

Learner Needs Analysis

The needs assessment survey was designed to capture the needs of learners who are interested in understanding the vocabulary and meanings of the ginans. It was important for this study to gather and understand the needs of the target population – members of the Ismaili community between 18 to 44 years of age living in Euro-American countries whose primary language is English – as their life experiences are different from those of their elders who migrated from their native land to Euro-American countries.

An online survey was developed to solicit the needs of the target population, which was administered between July 9, 2020, and September 10, 2020. As mentioned in the previous chapter, none of the questions in the survey was mandatory, which meant that the number of responses received varied from question to question.

Demographic Statistics. The survey attracted 515 participants from over 20 countries around the world (Table 4.5). Most of the respondents in the West were from Canada (49%), the United States of America (24%), and the United Kingdom (4%). Pakistan (9%) and India (5%) were the leading participants from South Asian countries.

Table 4.5

Survey Respondents by Country

Country	No. of Respondents	Percentage
Canada	237	46.0%
United States of America	116	22.5%
Pakistan	43	8.3%
India	26	5.0%
United Kingdom	21	4.1%
Other	72	14.0%
Total	515	100%

The age distribution of the respondents was grouped into seven intervals between 18 years and those over 75 years of age. With a 98% of response rate for the question on age, the highest number of respondents of the survey (24%) were in the age group of 55 to 64 years old (see Figure 4.5).
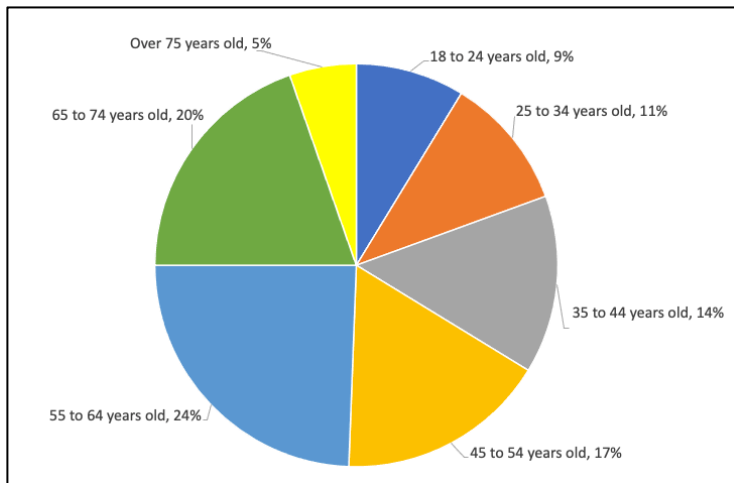
Figure 4.5. Survey Respondents by Age Groups

The question on gender was answered by 505 respondents, of whom 290 identified as female (57%), 213 as male (42%), and two as neither male nor female (0.4%). The lead by female respondents was consistent across all age groups except in the group aged 75 years and over, which was predominantly male at 70% (see Table 4.6).

Table 4.6

Survey Respondents by Gender

	Female	Male	Other	Total
18 to 24 years old	52%	45%	2%	100%
25 to 34 years old	52%	48%	0%	100%
35 to 44 years old	58%	42%	0%	100%
45 to 54 years old	64%	35%	1%	100%
55 to 64 years old	65%	35%	0%	100%
65 to 74 years old	56%	44%	0%	100%
Over 75 years old	30%	70%	0%	100%

In terms of educational attainment, over 75% of the respondents claimed to have at least one degree and only 3% of the respondents had not completed high school degree (see Figure 4.6). A further analysis of this data revealed that 89% of the respondents from South Asia had at least one degree as opposed to 75% of respondents in the diaspora with at least one degree.
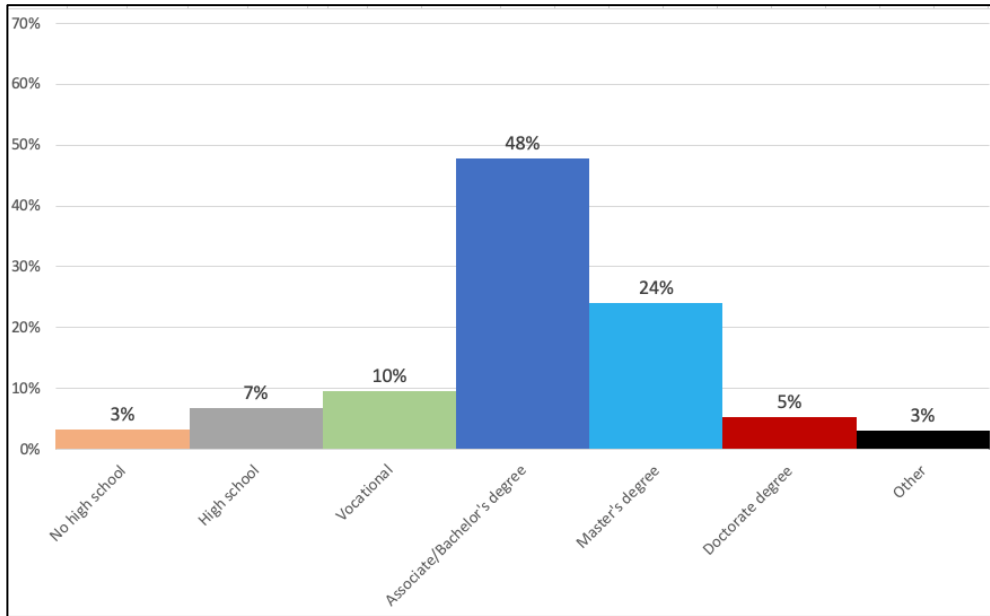
Figure 4.6. Survey Respondents by Highest Level of Education

The biggest group of survey respondents (37%) identified themselves as employed

professionals. Around 20% of the respondents identified as either business owners or self-

employed. Over one-quarter of the respondents (27%) identified as either retired or homemakers.

A visual summary of the primary occupation of the survey respondents is presented in the chart

below (see Figure 4.7).
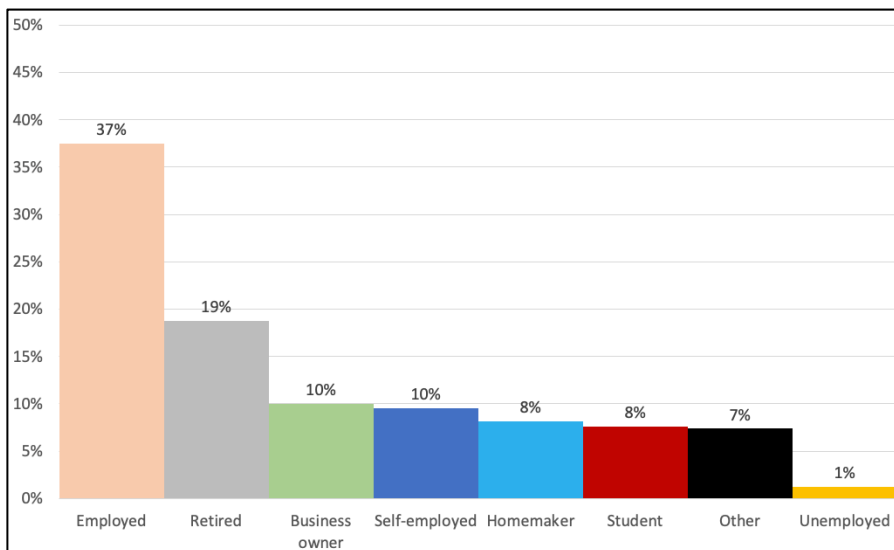
Figure 4.7. Survey Respondents by Primary Occupation

Identifying the target population. As evident from the demography analysis, there is a wide range of diversity in terms of geography, age, gender, education, and occupation among the respondents. Thus, owing to this geographic and demographic spread of the survey participants, the survey respondents were divided into two groups. The survey *target* group consisted of respondents who belonged to the target population define for this study – English-speaking Ismaili community members aged between 18 and 44 years who currently reside in Euro-American countries. The *general* respondents group comprised the respondents who fell outside the target population.

The target group respondents were identified by combining the responses of four specific questions in the survey (Q1 – Country of residence, Q2 – Age group, Q6 – Community membership, and Q10 – Primary language). The survey questionnaire is available in Appendix B. As depicted in the chart below (Figure 4.8), of the 515 total survey respondents, 496 respondents identified themselves as members of the Ismaili community. A total of 376 of these respondents live in Euro-American countries, and 238 of them designated English as their primary language. Finally, the pool of respondents in the target group was reduced to 71 when the age requirement was considered. Thus, the final size of the survey target was determined to be 71 or around 14% of the total number of respondents (n = 515).
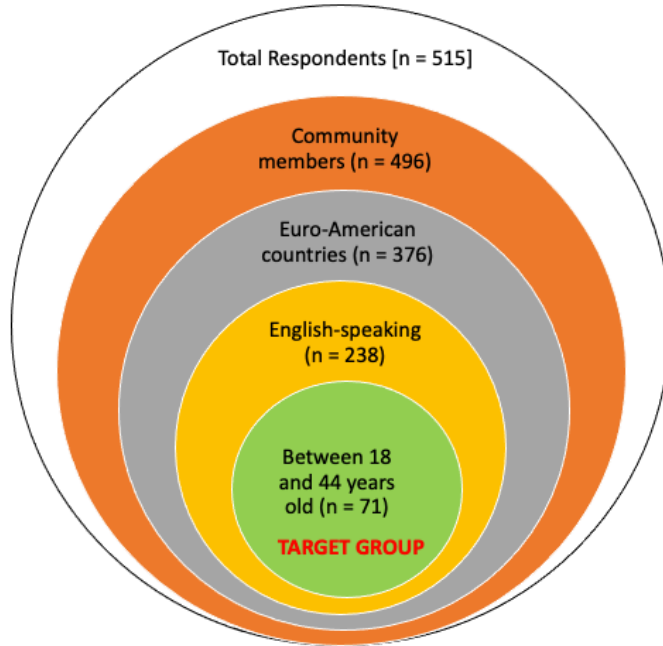
Figure 4.8. Identifying Respondents in the Survey Target Group

Needs of the survey Target Group versus General Group. It was helpful to benchmark the target group in relation to the other respondents to analyze the needs assessment survey data. Doing so helped in providing a consistent baseline in identifying, comparing, as well as contextualizing unique needs and attitudes between the two groups. For instance, when analyzing the data for the question on the importance of ginans in their lives (Q7), the expected difference in the attitudes between the two groups could now be visualized. A lower percentage of the target group (89%) attached moderate to high importance to ginans than the general community (97%). The evidence also validates the fundamental premise of this study that ginans must be made relevant to the younger generations based on their current socio-economic contexts by fostering understanding and meaning making. The use of information technology in learning and accessing ginans is important of this premise for the target group.
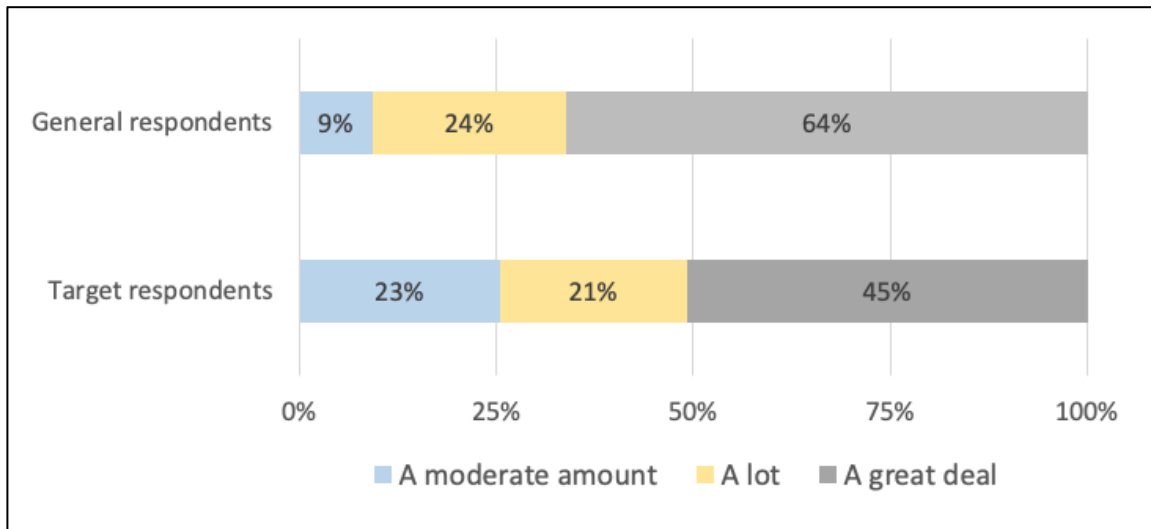
Figure 4.9. Importance of Ginan in Target and General Respondents Groups

A strong agreement between the needs of the two groups was observed with regards to having access to ginan resources. An overwhelming majority (97%) in both groups attached moderate to high importance to having online access to ginan resources. On the question of preferred devices to access online ginan resources as well, there was notable synergy in the needs of the two groups. The use of mobile phones remained the most preferred device in the target group for 97% of the respondents in comparison to 91% for other respondents. The two groups diverge in their preferences, however, when it comes to accessing ginan resources in the form of books, CDs, cassettes, etc. More than one-quarter (28%) of the respondents outside the target group attach moderate to high importance to such analog resources in comparison to less than one-tenth of the respondents in the target group.

Given the demography of the target group, it is not surprising to find an overwhelming demand (97%) for the Latin script and English as the language of instruction for the ginans. It was surprising, however, to find the English language to be preferred by 91% of the respondents in the general group. While none of the other scripts come close to the strong support shown for

74

English, Khojki and Gujarati remain alive and important in the community even today (see Figure 4.10). The findings of the survey reflect the historical decline of Khojki and Gujarati scripts in favour of English as a substantial number of the community members have moved away from the Indian subcontinent to Euro-American countries across the globe.
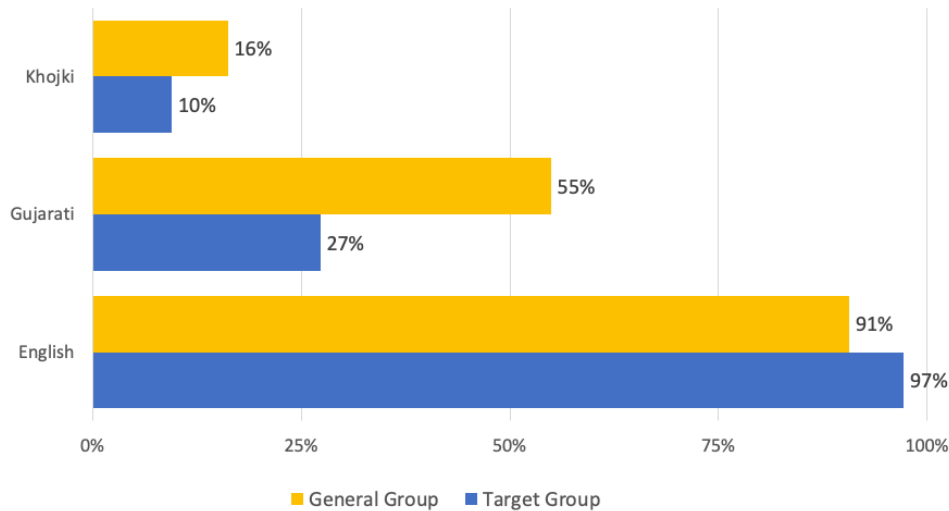
Figure 4.10. Language Preference for Learning and Understanding Ginans

To ascertain the online content types most needed by ginan learners, a variety of options were presented to the respondents to rank as part of the needs assessment survey. These options were ranked independent of each other by the respondents. Having access to ginan texts and translations in English was ranked as the most desirable resource for learning and understanding of Ginans (Figure 4.11). This outcome is not surprising as it is a common practice to use ginan text and translation side-by-side during in-person instructional sessions.
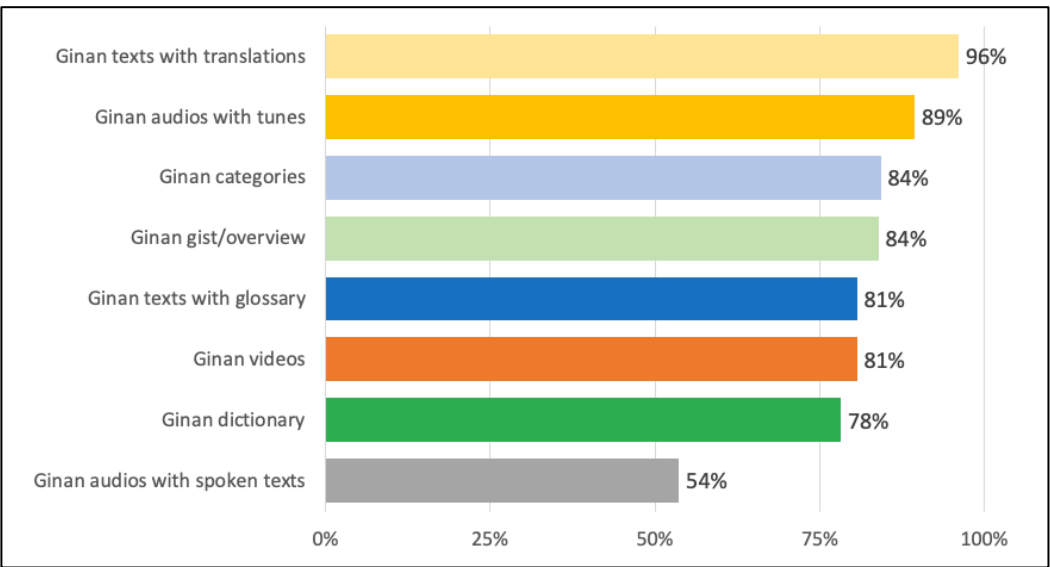
Figure 4.11. Useful Resources for Learning and Understanding Ginans

The second most desired content type was audio recitations of ginans. In an in-person instructional setting, the instructor is responsible for reciting ginans to the learners. In an online and self-learning setting, where there are no instructors, the availability of digital audio is crucial. The need for information on ginan categories is also ranked as desirable by 84% of the respondents. The ginan categories are used to group ginans based on various ceremonial and topical themes. While the practice varies from country to country, some ginan publications organize and list ginans based on such categories. For instance, the table of contents from a 2010 ginan book published in Kenya shown in the figure below (Figure 4.12), lists several such categories of ginans. In recent years, the knowledge of ginan categories has been confined to the community elders and experts who may choose to impart this knowledge to learners during their in-person instructional sessions.

```
                  List of Ginans
                    (Cateogirc)

No.          Title                    Page No.

Bandagi
1.    AAD THAKI EK SUN NIPAAYA          1
25.   EK SHABAD SOOÑO MERE BHAI        46
31.   HARDAM KARO ABHIYAAS             55
38.   JOGI SO JO JUGA JUG JOGI         66
39.   JOGKA PENDA DOHELA               67
58.   PRATHAM JOG KERA MUL             99

Bandagi/General
45.   KESARI SI(N)H SAROOP BHULAAYO    78
48.   MAARI AATAMNA ODHAAR ( SAKHI )   82
49.   MAHAPAD KERI VAAT ( SAKHI )      83

Chogadia (Evening)
10.   ASHAJI SAACHO TU ALAKH           15
40.   JOOGME(N) PHIRE SHAHJI MOONERI   69

Chogadia (Morning)
80.   SREVO SREVO MOMAN ALI EK        143

Chogadia (Morning & Evening)
13.   BHAI TINI VIREJIU(N)             21



                      xvi
```

Figure 4.12. Sample of Ginan Categories[22]

Another sought-after resource for learning and understanding ginans that is ranked considerably higher in the needs assessment survey was the summary of individual ginans, commonly referred to as ginan "gist" in the community. Over the past decade, it has become a common practice for reciters to read out the gist of the ginan in English that they are called upon to recite during congregational services held for special occasions. The gist texts explain the message and sentiments of ginans in broad strokes for the English-speaking members of the congregation who often struggle to understand what is being recited. Despite this being a

---

[22] Source: Ginan-e-Shariff: Our Wonderful Tradition Vol II, Kenya, 2010

common practice, the community institutions have yet to produce any publication with ginan gists that can be readily accessed by the community.

Ginan dictionary and multimedia ginan videos were ranked equally high in the survey. When analyzed based on specific groups, only 67% of the respondents in the target group attach moderate to high importance to a ginan dictionary as opposed to 80% of those outside the target group. The need for multimedia videos is also relatively less pronounced in the target group at 77% as opposed to 81% in the general group. The least desired ginan resource turned out to be the audio recording of spoken (as opposed to sung) words or lyrics of ginans. It was a common practice in the past for community elders and instructors to read aloud lyrics of ginans to aid their students to learn proper pronunciations of the words used in ginans.

Final Prototype Design

For transforming the initial design into a functional prototype, both the needs of the learners as well as the best practices of community instructions from experts were synthesized and ranked to identify the feasibility of the desired features.

Table 4.7

Overall Feasibility of Features for the Prototype

	Feature Name	Feature Category	Content Availability	Time Availability	Technical Feasibility	Overall Feasibility
1	Ginan text	Content	Yes	Yes	Yes	Yes
2	Ginan translation	Content	Yes	Yes	Yes	Yes
3	Prescribed tune	Content	Yes	Yes	Yes	Yes
4	Glossary	Content	Yes	Yes	Yes	Yes
5	Semantic profile	Content	Yes	Yes	Yes	Yes

	Feature Name	Feature Category	Content Availability	Time Availability	Technical Feasibility	Overall Feasibility
6	Gists of the ginans	Content	No	Yes	Yes	No
7	Composer identification	Content	Yes	Yes	Yes	Yes
8	Ginan videos	Content	No	No	Yes	No
9	Ginan categories	Interface	Yes	Yes	Yes	Yes
10	Ability to markup text	Interface	N/A	No	No	No
11	Access to other resources	Interface	Yes	Yes	Yes	Yes
12	Ginan uncommonness	Interface	Yes	Yes	Yes	Yes
13	Extended audio player	Interface	Yes	Yes	Yes	Yes
14	Ask-an- Instructor option	Interface	No	No	Yes	No
15	Ability to print	Navigation	Yes	Yes	Yes	Yes
16	Pronunciation guide	Navigation	Yes	Yes	Yes	Yes
17	Ability to download	Navigation	Yes	Yes	Yes	Yes

Ranking and feasibility analysis of features. The list of candidate features was compiled based on analyzing the data gathered from the learner needs assessment survey as well as reviews from the community experts. The list comprised 17 features reflecting both the needs of the target group and the best practices of the experts. The features were weighted and ranked based on the number of learners and reviewers supporting them. Each feature underwent feasibility analysis for inclusion in the prototype. Table 4.7 summarizes the results of the feasibility analysis of the ranked features.

Implementing the prototype. Of the seventeen features enumerated for the prototype, five features were already part of the initial design. The top four ranked features were independently and unanimously validated by all three experts. These four features were also

79

designated as highly desired by the target population. The semantic profile feature ranked fifth in the list also received high acceptance from the experts for its integration with other resources and versatility as an online resource for learners and instructors alike. The feature that was ranked sixth did not pass the feasibility test as ginan gists are not readily available in the wider community for inclusion in the prototype.

The feasibility analysis of adding the feature of ginan videos posed an interesting dilemma. All the three expert instructors unanimously rejected the idea of incorporating ginan videos as a pedagogical tool, yet there was overwhelming support for this feature by the learners (93%). At the root of this dilemma was the fundamental question of who gets to decide what resources are best suited for teaching ginans. In most educational settings, the decisions about curriculum as well resources are made by educators, not students. With this historical and logical precedent, the feature to include ginan videos was rejected for inclusion in the prototype. This decision was further supported by the unavailability of time required to develop such a resource for the selected ginans for this study. It is, however, important to point out that the technology to integrate online ginan videos is available and could be incorporated in the prototype design if and when desired by the community learners.

Two additional features that were deemed infeasible included an online tool to markup ginan texts and the Ask-an-Instructor feature. Although both these features are creative and may prove to be effective in the learning and understanding of the ginans in an online environment, they remain infeasible for this research due to the underlying technical infrastructure as well as financial and the time needed for their development and inclusion for the prototype.

In summary, of the seventeen ranked features for the implementation of the prototype, five features were already part of the initial design (29%), four features failed the feasibility test (24%), and eight new features (47%) were included in the final prototype design. The following screenshot (Figure 4.13) and table (Table 4.8) are helpful in visually identifying and illustrating the eight new features (labelled A through H) implemented in the porotype are.

Figure 4.13. New Features Incorporated in the Prototype Design

Table 4.8

New Features Implemented in the Prototype

	Feature Type
A	Composer information
B	Ginan categories
C	Access to other resources
D	Ginan uncommonness

	Feature Type
E	Extended audio player
F	Ability to print
G	Link to the pronunciation guide
H	Ability to download

Learner Validation Analysis

From the 71 respondents identified to be part of the target population, 34 respondents (47%) volunteered to be contacted for any follow-up studies. Upon contact, a total of 11 volunteers (32%) sent in their responses. The smartphone was the most popular device (46%) used by the respondents to access the prototype, followed by laptop (27%) and desktop (18%) computers. On average, most of the respondents (73%) spent between 5 and 30 minutes using the prototype (see Figure 4.14).

On the 5-point Likert scale, the overall rating of the prototype was 4.64 (93%) based on the weighted average of the responses. The ratings of the three review categories (content, interface, and navigation) also received ratings of between 3.91 to 4.64 based on weighted average as summarized in the table below (see Table 4.9). Most of the respondents (91%) indicated that they would make use of the proposed online ginan tool designed and developed as part of this study.
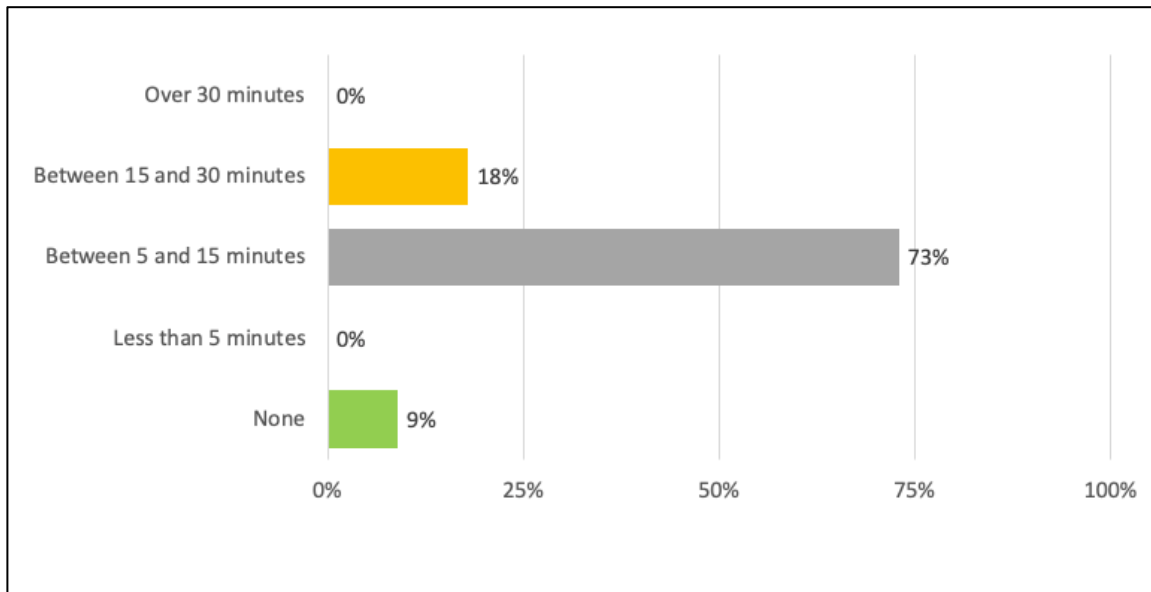
Figure 4.14. Average Time Spent on Exploring the Prototype

Table 4.9

Validation Score of the Prototype Category Based on Weighted Average

	Average Score (Max 5.00)	Percentage
Content	4.64	93%
Interface	3.91	78%
Navigation	4.27	85%
Overall	4.82	96%

The comments received from the participants were also reflective of the overall acceptance and validation of the prototype. One of the participants commented that they "really enjoy the vast variety of options that come with the tool such as the tune, text, translation, glossary, and the pronunciation guide. It makes the tool extremely useful!" Another respondent

articulated their validation of the prototype by stating: "Awesome work on this, I genuinely believe it will benefit many ginan learners - myself included" (Respondent #8).

A few respondents also pointed out some shortcomings of the interface. For instance, a respondent noted that the "[b]uttons on top which says download pdf, text, the print were not working" (Respondent #1). Another respondent suggested developing the tool as "an app instead of rendering webpages on the mobile devices" (Respondent #4). In general, the results of the learner validation were very favourable toward the design, efficacy, and acceptance of the prototype. The following comment from one of the respondents aptly summarizes the sentiment and utility of the tool:

> I appreciate the various components of this tool! Being a religious educator and someone who is looking to learn new ginans, I appreciate the meanings of the verses and the words that I may not be familiar with. The audio is also very helpful. I also appreciate the glossary for easy access and the sources at the bottom of the ginan because it helps to provide a context of the chosen ginan…. Thank you for this and I look forward to seeing this site when it is complete. (Respondent #11)

CHAPTER 5: DISCUSSION

This chapter briefly summarizes the activities undertaken in prototyping an online ginan learning tool. It provides answers to the research questions put forth for this research. In addition, this chapter discusses the limitations of this research followed by its implications for future studies.

Summary of Activities

This study prototyped an online ginan learning tool for English-speaking learners interested in understanding the vocabulary of the ginans. As part of this research, a design-based approach was taken to systematically gather and analyze the needs of learners to prototype an online tool for learning and understanding the vocabulary of the ginans. Using two iterations of design and evaluation, the initial design of the tool evolved to a functional prototype that was validated by experts and learners.

In the initial design iteration, the theoretical conjectures of language learning were blended with traditional methods and materials of teaching ginans. Based on the literature review, *glossing* was identified as a crucial best practice for vocabulary semantization for L2 learners. Online glossing, which is the practice of embedding brief definitions in L1 for content words in L2 texts using web links, is dependent on the availability of lexicons such as glossaries and dictionaries. To this end, a technology-mediated semantic analysis of a sample ginan corpus was undertaken to create a ginan lexicon. This lexicon was used for the online glossing feature of the prototype.

The initial design of the prototype served as a starting point for experts and educators to critique and review the design for efficacy in the context of ginans. The feedback from the

reviews was translated into design features that were analyzed for feasibility in terms of content availability, technical complexity, and implementation time. Thirteen of the seventeen features that passed the feasibility criteria were chosen for the final design iteration.

In the final design iteration, the selected design features from the initial design iteration were reconciled with the needs of the learners captured through an online survey. While the survey was open to the public and attracted over 500 responses from across the globe, there were 71 respondents in the target population defined for this study: English-speaking Ismaili community members between 18 and 44 years of age living in Euro-American societies. The survey validated the importance of ginans that the target group attached to the tradition of ginan and its appetite for more online resources to make ginans relevant to the contemporary lives of the diaspora community.

As part of the final design iteration, volunteer respondents from the target group were invited to evaluate the prototype for its efficacy and ease of use. The results showed that the prototype developed in this study was accepted with a high degree of satisfaction in meeting their needs. Moreover, the design elements of the prototype, including content, interface and navigation were tested and found to be highly intuitive and effective by the volunteer respondents.

Conclusions

The learning of the ginans by successive generations is essential if the tradition and the teachings of ginans are to survive in Euro-American societies. The purpose of this study was to investigate, design, and validate an online tool for learning and understanding the vocabulary of

the ginans. This study used a design-based research (DBR) methodology to prototype the tool. The research questions that guided this study are discussed next.

Research Questions

The purpose of this research was to prototype an online Ginan learning tool guided by an overarching research question: *How can an online tool be designed to facilitate the learning and understanding of the ginans by English-speaking learners?* While the open nature of the research question allowed for prototype design to be iterative and evolutionary, the three sub-questions ensured a pragmatic and realistic outcome.

The sub-questions for this research were developed to solicit perspectives from multiple disciplines to inform the design choices of the final prototype. The first research sub-question synthesized best practices of library science and corpus linguistics in organizing and analyzing available ginan materials for this research. The focus of the second sub-question was anchored in education to gather and analyze learners' needs in the context of learning and understanding ginans. Finally, the third research sub-question incorporated perspectives from educational design and technology and computer science in the design and development of the prototype.

RQ1: How can the existing ginan corpus be curated to support online language learning from a digital humanities perspective?

Taking a digital humanities (DH) perspective in addressing a research problem requires openness to a multidisciplinary approach and commitment to using information technology to find solutions. While anchored in L2 semantization in applied linguistics, this research also adopted best practices from two additional practice-based disciplines: library information science (LIS) and corpus linguistics (CL). Both these disciplines have unique approaches when it comes

to semantic analysis. Perhaps the analogy of a telescope can be used for how semantic analysis is used in LIS to orient artistic and literary *works* to the *bibliographic universe* i.e., the wider hierarchy of existing knowledge. In CL, however, semantic analysis is used to undertake microscopic studies of *textual works* to build lexicons.

In addition to using computing for corpus processing and semantic analysis, the use of information technology also became indispensable for digitally curating ginan materials for online semantization. Much like DH, digital curation has been at the forefront of interdisciplinary collaboration. Digital curation can be described as one of the methods through which "digital humanities focuses on a range of work products such as scholarly editions, text corpora, marked-up text, thematic research collections, previously annotated or analyzed data, and finding aids or bibliographies" (Poole, 2017, p. 1172).

As discussed in Chapter 2, the notion of "digital text" in DH to include digital media makes it easier to combine best practices of LIS and CL for digitally curating available ginan materials to facilitate online semantization. In addition to embedding online glosses in the ginan texts, the semantization prototype also provided links to various other ginan materials such as manuscripts, non-English texts, recitals, and academic studies. The overall process followed to make these materials accessible for the online semantization prototype is illustrated in the figure below (Figure 5.1).
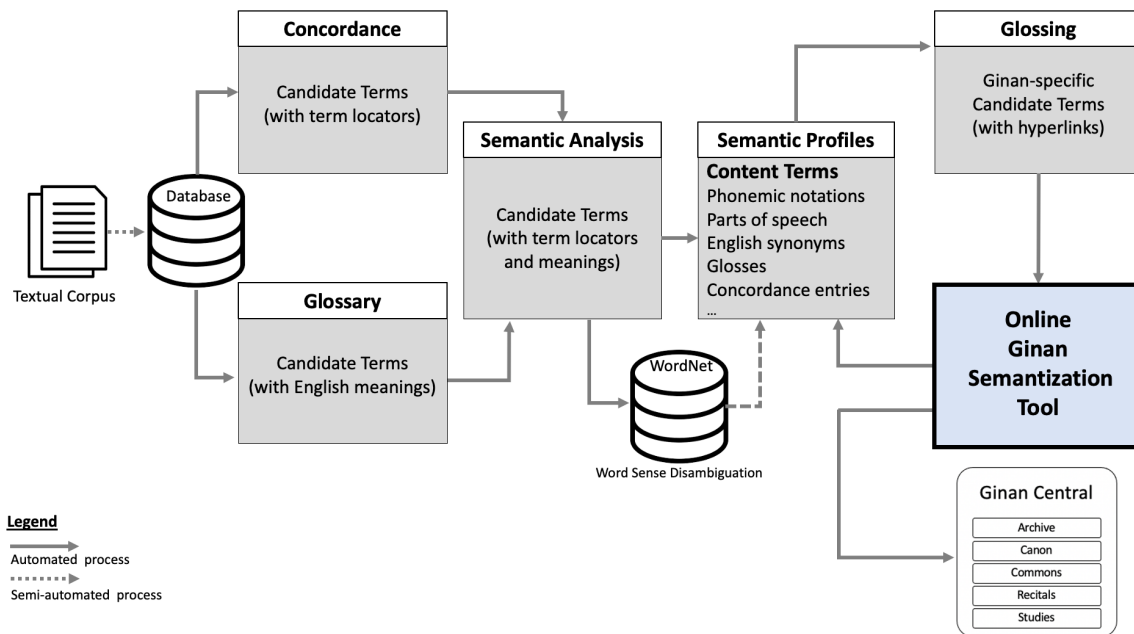
Figure 5.1. Process Overview of the Prototype

As discussed in Chapter 3, the concordance and glossary data sets were stored and matched in the database to identify pertinent ginan terms for the semantization prototype. The creation of an online lexicon for each of the terms was developed based on the semantic analysis process outlined in previous chapters. As was discussed in Chapter 4, the word sense disambiguation process was semi-automated as it required manual verification of the match results from the automated algorithm. The ginan texts used in the online prototype were further processed to embed glosses as hyperlinks that linked content words to their semantic profile. The glossing process used is discussed under RQ3 below.

The online learning tool also provided ginan-specific links resources from five additional ginan collections available from the Ginan Central portal. These collections include *Ginan Archive* for primary sources of ginans, *Ginan Canon* for non-English printed texts of ginans, *Ginan Commons* for additional English texts of ginans from community members, *Ginan*

Recitals for accessing digital audio recordings, and *Ginan Studies* for English translations published in academic studies. The seamless integration of the online prototype with Ginan Central collections was facilitated by ingesting a unique Ginan Master identifier assigned to each ginan in the Ginan Master Index registry available at Ginan Central.[23]

RQ2: What are the needs of the community members interested in learning and understanding the ginans?

The needs of the target population to inform the design of the online ginan learning tool have already been discussed in Chapter 4. Based on the analysis of the needs assessment survey data, several conclusions on the broader needs and aspirations of the community about the ginans can also be made based on the aggregated data collected for all the survey participants.

Community Profile. A typical community member is an educated citizen of an English-speaking country such as Canada. They consider ginans to be an important part of their life even though the language of ginans remains mostly foreign to them. They value the emotive and performative aspects of the tradition that help them express their devotion and solidarity to the Ismaili faith and community. Although their exposure to one or more Indic languages at home and in the community allows them to understand a few ginan terms, they remain very concerned that the ginans, and more so the teachings that the tradition encapsulates, will be lost if nothing is done about it by the community and its institutions. Perhaps this concern is why their interest in understanding the meanings of ginans remains heightened.

---

[23] The Ginan Master Index (GMI) is an evidence-based master registry of Ginans available from Ginan Central at https://ginans.usask.ca/master.

Community Needs. When it comes to the needs of the community in the context of learning and understanding ginans, the availability of ginan materials in English and the establishment of curriculum-based ginan classes remain two dominant themes.

Availability of Ginan materials. The broader thematic need of ginan materials can further be categorized into three sub-themes: content types, content sources, and online accessibility. As depicted in the table below (Table 5.1), the most desired online content types in the community remain ginan texts, translations, and recitals. Several community members, however, have also indicated their preference for using printed ginan books with English texts and translations during congregational services.

Table 5.1

Most Desired Ginan Content Types Based on Weighted Average

	Average Score (Max 4.00)	Percentage
Ginan texts with translations	3.40	85%
Ginan audios the with tunes (recitals)	3.09	77%
Ginan categories	2.80	70%
Ginan gist/overview	2.79	70%
Ginan videos with text, translation, audio, etc.	2.76	69%
Ginan texts with links to glossary	2.73	68%
Ginan dictionary	2.62	66%
Ginan audios without tunes (spoken)	1.87	47%
Other	0.92	23%

It is important to note that the Ismaili Tariqah and Religious Education Board (ITREB) has been historically responsible for making educational materials available to the community members, including the ginan materials. Today, the Institute of Ismaili Studies (IIS) is also responsible for training interested and qualified community members to teach its curriculum in the community through its Secondary Teacher Education Programme (STEP) program.[24] An unfortunate consequence of this partnership between ITREBs and IIS has been the unavailability of new ginan materials for the community. In a recent reflective article, a former Co-Director of the IIS noted that (Karim, 2021):

> Whereas substantial work has been carried out in examining Arabic and Persian documents, the study of Indic manuscripts (bearing content such as Ginans) has been minuscule in the last four decades. Research is also conducted on the transnational Ismaili community's living traditions, but it is not published for the most part. The Institute prioritizes a rationalistic and civilization-centred approach over faith perspectives in its course instruction and religious education curricula for the global Jamat [community]. These characteristics, viewed as appropriate for the IIS's particular mandate, have, however, raised an air of ambivalence that has apparently produced unintended consequences. … Furthermore, donors who have contributed substantially to the endowment are perplexed by the asymmetry in Ismaili areas of research. (para. 4)

The divergence of perspectives on the ginans between Satpanth Ismailis and certain Ismaili institutional circles remains a barrier in making any significant headway in preserving the ginans

---

[24] More information on the IIS STEP program is available at https://www.iis.ac.uk/graduate-studies/step.

through formalized teaching and transmission. While these Satpanth Ismailis continue to revere

ginans, the "Hinduistic" elements of the ginans remain problematic for the institutions.

Consequently, community members use various external websites to access ginan

materials, even when they might regard these sites to be "unofficial" sources with potentially

"unauthentic" content (Figure 5.2). The lower usage of the IIS website is indicative of nominal

materials on ginans. Thus, there is a deep desire in the community to use online ginan websites

and content that are either produced or endorsed by community institutions. As one community

participant stated, the need for developing "a unified website which is accredited by IIS that is

made available globally and it should have authentic text and raags of ginans" (Respondent

#506).
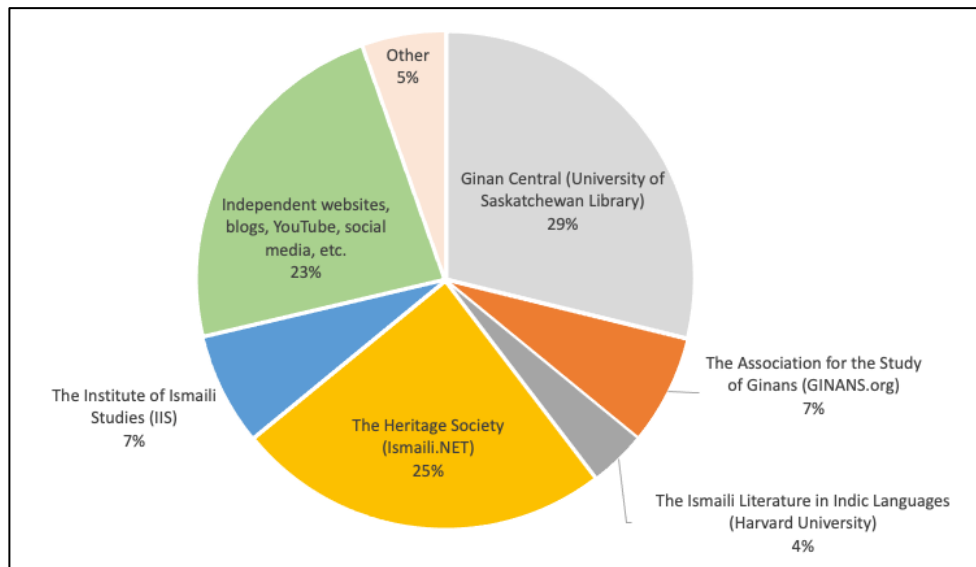
Figure 5.2. Most Preferred Websites for Accessing Ginan Materials

Ginan Classes. The second thematic need that surfaced during the analysis was the

establishment of the ginan classes. Currently, ginans are taught and transmitted by community

volunteers at their initiatives. While volunteering to teach ginans is an admirable trend in the

community, there is no standard or structured way of tracking which ginans are taught and how locally or nationally. There is a deep desire in the community for institutionalized development and administration of curriculum-based in-person and online ginan classes. The expectation here is that the curriculum for these classes will not only teach the meanings and tunes but will ensure that the history and teachings of ginans are also made relevant to the community's contemporary context in the diaspora.

In a broader context, this need of the community to seek the relevance of ginans in contemporary contexts is indicative of the community's search for new and equivalent cultural expressions of its faith and devotion in English-speaking and Christian societies. In this context "translation" is no longer an exercise in finding linguistic equivalence but becomes a quest to seek equivalence of one's faith in the local culture. Contextualizing such meaning-making from a translation theory perspective, Stewart notes (2006, p. 286-87):

> "[T]he search for equivalence in the encounter of religions—when understood through the translation models we have characterized as literal, refractive, dynamic, and metaphoric—is an attempt to be understood, to make oneself understood in a language not always one's own; it does not necessarily reflect religious capitulation or theological ignorance or serve as the sign of a weak religious identity…. The texts that reveal actors attempting to locate commensurate analogues within the language tradition capture a unique 'moment' in the process of cultural and religious encounter, as each tradition explores the other and tries to make itself understood."

The institutions such as IIS and national ITREBs have a responsibility to make the tradition and the teachings of the ginans relevant to the contemporary contexts of the Ismaili community rather

than wish the tradition away. A key component of fulfilling this responsibility will be the reformulation of the curriculum of the ginan classes in the diaspora to meet the needs and expectations of the community as evident from the needs assessment survey conducted as part of this study.

Another important component in reformulating the ginans classes, also evident from the survey results, is the use of information technology. This where the use of the online ginan tool prototyped in this study can be useful. The major implications of doing so is that the online ginan learning tool: 1) meets the personal needs and cultural expectations of the English-speaking learners in diaspora, 2) presents an avenue for the development of new curricula and new approaches to teaching and learning, 3) provides a structure for teachers and learners to improve and enhance learning outcomes, 4) provides motivation and flexibility for learners to remain engaged, 5) provides readily accessible online resources specifically designed for learners to complement traditional and pedagogical best practices, and 6) ensures revitalization and preservation of traditional language and heritage.

The anticipation and acceptance of the online ginan learning tool based on the validation survey is encouraging for the long-term preservation of the tradition and teaching of ginans. The communnity's receptivity of the tool, however, can be greatly enhanced through the participation and endorsement of the community institutions, which remain responsible for the religious education of the Ismaili community.

RQ3: In what ways can information technology be useful in learning the language of ginans?

Notwithstanding the lack of standardized romanization convention for transliteration of ginans among the community and academia, the use of information technology is crucial for

enabling glossing in ginan texts. The process of glossing required going through all the documents in the selected corpus and creating hyperlinks for all matched occurrences of content terms to link them to semantic profiles. This process is tedious and prone to errors if done manually.

To ensure efficient and accurate results, the process of glossing for this research was automated. Much like the semantic analysis process discussed in Chapter 3, the automated glossing process was also designed to be extensible and replicable. The extensibility of the process ensures that the underlying corpus could be extended to induct additional ginan texts and lexicons without having to rewrite the program. The automation of the process also guarantees that the results are reproducible when repeated over time.

The glossing process. The glossing process used in this research utilized the online lexicon, which was developed by integrating concordance and glossary of the ginans. In today's age of full-text searching, where entire documents can be indexed efficiently and economically for full-text indexing, the need for a concordance may seem questionable. While full-text indexing may be efficient for searching terms in individual documents, it is unable to link documents together semantically. As seen in this research, the ability to build semantic relationships amongst resources is crucial for developing semantic profiles to enhance semantization. As noted in Chapter 2, both concordance and glossary-based glosses remain essential and effective for L2 learners.

In this study, the concordance served the dual purpose of identifying candidate terms for semantization and collocating the terms within the selected corpus. The candidate terms obtained from the concordance were matched against the terms in the glossary to identify content terms

for semantization. This process of identifying content terms assumes that terms included in the glossary are deemed pertinent for a learner to know and understand. Since glossaries typically do not provide the location of terms in the corpus, the matching of concordance and glossary terms provides the additional benefit of generating glossaries for specific ginans.

Once the glossary terms for a specific ginan are identified, the glossing algorithm is programmed to go through each line of the ginan text and insert hyperlinks for terms in the line that matches any of the candidate terms. At the end of the process, all candidate terms appear as hyperlinks in the ginan text that link to their respective terms in the ginan-specific glossary generated at the end of the text. Additional links to pronunciation sound files and semantic profiles are also incorporated into the ginan glossary.

A general conclusion that can be drawn from this study is that the use of information technology can complement and enhance the transmission of traditional knowledge. From a learner-centric perspective, the use of information technology is imperative for engaging learners to embrace traditional languages and traditions.

Discussion

This study prototyped an online ginan learning tool for English-speaking learners interested in understanding the vocabulary of the ginans. The methodology as well as various analyses used in this study ensured and verified that the tool was not only developed based on the needs of the community but were also accepted by learners as an effective tool. The design-based approach and the use of open-source technology allow the tool to be used by local and regional community institutions to complement the ginan curriculum and instructions.

This study has particularly highlighted the survey responses of the English-speaking Ismaili community members aged between 18 and 44 years who currently reside in Euro-American countries, including the United States, Canada, and the United Kingdom. These results are crucial for meeting the needs and expectations of the community members who wish to learn and understand the teachings of ginans and remain responsible for carrying them forward. These findings will be useful in charting the future direction of the relevant community institutions, including the IIS and national ITREBs.

Overall, the survey found that the availability of ginan resources in English remains a crucial need for the community members who have little or no knowledge of the language of the ginans. The development and dissemination of curriculum-based educational programs and supporting materials for the ginans emerged as the most urgent and unmet expectations of the community. Additionally, the survey results validated the community's longstanding desire for a more pronounced acknowledgement of the heritage of ginans and its significance in the communal and scholarly undertakings of its institutions and leaders.

Research Outcomes

The research outcomes of this design-based research are made available on the companion website developed for this research (Figure 5.3). These outcomes include the semantic profiles of content terms, the initial design of the online ginan learning tool, and the functional prototype of the online ginan learning tool. The thesis companion website can be accessed at: https://ginans.usask.ca/semantics/
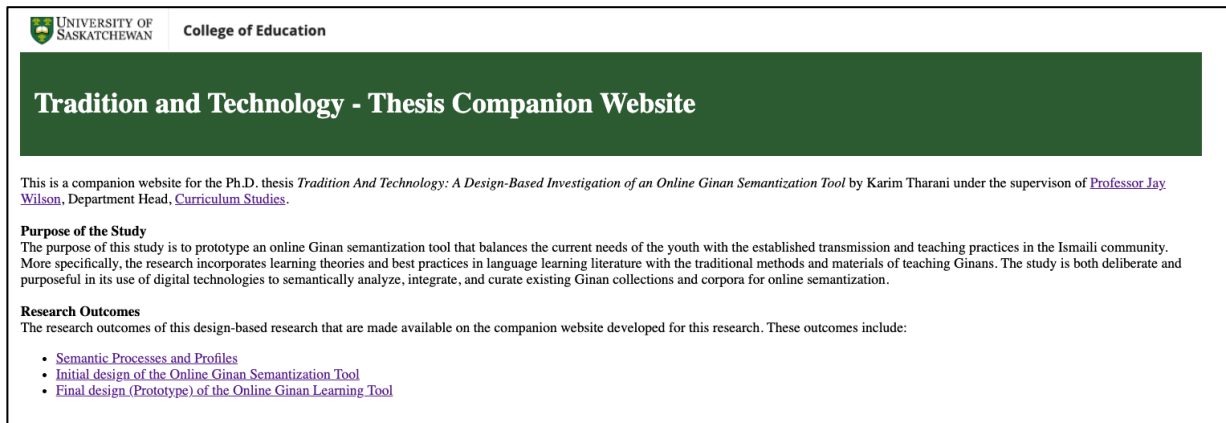
Figure 5.3. Thesis Companion Website

Semantic Profile

The semantic profile template proposed in this thesis was implemented and populated by semantically analyzing 143 content terms (Figure 5.4). These terms along with their profiles can be accessed through the thesis companion website (Figures 5.4 to 5.6) or directly at:

https://ginans.usask.ca/semantics/analysis/
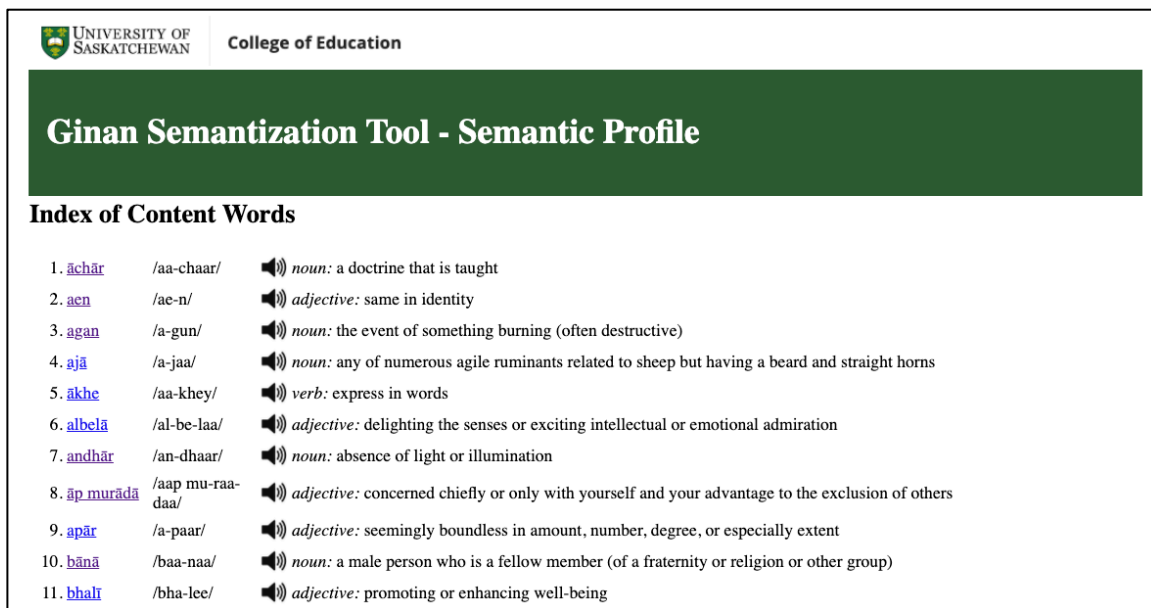
Figure 5.4. Sample Content Words

Figure 5.5. Online Semantic Profile – Part I

Figure 5.6. Online Semantic Profile – Part II

Initial Prototype Design

The initial design of the prototype was designed by applying best practices from the L2

semantization literature to the available ginan materials. This design was also used to solicit

feedback from the community ginan experts and teachers. The initial design can be accessed

through the thesis companion website (Figures 5.7 to 5.9) or directly at:

https://ginans.usask.ca/semantics/poc/

Figure 5.7. Initial Design of the Prototype – Part I

Figure 5.8. Initial Design of the Prototype – Part II

Glossary

ajā: any of numerous agile ruminants related to sheep but having a beard and straight horns
bharam: an erroneous mental representation
bhav: the essential qualities or characteristics by which something is recognized
gāli: the state of the world as it really is rather than as you might want it to be
girabh: take hold of so as to seize or restrain or stop the motion of
jīv: the immaterial part of a person; the actuating cause of an individual life
jīvan: the course of existence of an individual; the actions and events that occur in living
kesrī siṅh: large gregarious predatory feline of Africa and India having a tawny coat with a shaggy mane in the male
ochinto: happening unexpectedly
pāḷo: act in accordance with someone 's rules, commands, or wishes
sang: in contact with each other or in proximity
sarūp: your consciousness of your own identity
shūdh: having no faults; sinless
sove: a natural and periodic state of rest during which consciousness of the world is suspended
ṭāḷo: avoid and stay away from deliberately; stay clear of
vachan: the power or authority to command
vāro: give up with the intent of never claiming again

Source(s):

- Alwaez Kamaluddin, Muhammad. (2011). *50 Ginans with English Translation and Glossary - Volume 1(Second edition)*.
- Alwaez Kamaluddin, Muhammad. (1990). *Concordance of the Ginans - Volume 1*.

Figure 5.9. Initial Design of the Prototype – Part III

Final Prototype Design

The final prototype design represents the final design of the tool with features recommended by ginan experts as well as community ginan learners. The design was also validated by the target users. The prototype design can be accessed through the thesis companion website (Figures 5.10 to 5.12) or directly at: https://ginans.usask.ca/semantics/prototype/

Figure 5.10. Final Design of the Prototype – Part I

Figure 5.11. Final Prototype Design – Part II

Figure 5.12. Final Design of the Prototype – Part III

Limitations of the Study

The study frames the learning of the ginans to be equivalent to the learning of a foreign language by English-speaking online learners. Furthermore, the online prototype is designed to meet the immediate needs of the Ismaili community. It is important to recognize, however, that while the prototype may be of limited use outside the community, the methodology and findings of this research can be beneficial to other communities who find themselves in a similar situation.

The use of variant romanization conventions in the available ginan materials made it difficult to synthesize and integrate diverse ginan materials purposefully. Consequently, the corpus used in this research was limited by its selection of materials with compatible romanization conventions.

Recommendations for Future Studies

The focus of this study has been on semantization of the ginan vocabulary. With its deliberate use of information technology, this study sought to incorporate the Western best practices in language learning to complement traditional ways of transmitting and teaching ginans. Taking this approach has surfaced two critical areas of need that can benefit from further innovative experimentation and academic research. These are: 1) the need for standardized romanization convention for ginan texts, and 2) the availability of a ginan dictionary.

Standardized Romanization Convention. It is important to recall that while the ginans were canonized using the Khojki script, the language of ginans itself is not Khojki, but a mixed language that draws from multiple Indic and non-Latin languages. The phonetic demands of the oral and polyglottic nature of the ginan language posed challenges for the limited phonetic strength of the Khojki script. Unfortunately, these shortcomings were never addressed by the community systematically to standardize Khojki. As a result, certain inaccuracies in the transcribed texts were passed on as the ginan corpus was canonized from Khojki to a more established Gujarati script. The canonized ginan corpus in Khojki and Gujarati were then used as the basis for the romanization of ginans into the Latin or English writing system.

Variant romanization conventions. With the dispersion of the community from South Asia, the local community leaders and elders were left to their own devices for developing romanization conventions for the ginans. As a result, variant ginan romanization conventions started to emerge from different countries where the community resided. In the academic circles as well, there are variant romanization conventions in use for ginans. For instance, the sound 'sh' (as in the word shore) is found to be romanized with 'sh' (Asani, 1992) and ś (Shackle & Moir,

2000). The academic romanization conventions, however, are distinct from their communal counterparts in their rigour, compactness, and attention to the underlying etymology of the vocabulary. Unfortunately, due to the dispersed nature of the community as well as limited attention or action by the responsible community institutions such as national ITREBs and IIS, the issue of standardizing romanization conventions, both among and across communal and academic circles, remains unattended.

Phonetic (sound-based) convention. One approach to overcome this impasse of standardizing variant romanization conventions may be to move away from the existing practice of text-based romanization conventions. These text-based or *orthographic* romanization conventions remain deficient in representing the multilingual phonetic richness of the language used in the ginans. The use of sound-based or *phonetic* transcription to focus on the oral/aural properties of the ginan language may be a better approach for the romanization of the ginans. With this approach, the content word such as āchār could be assigned the phonetic notation of /aa-chaar/. By ensuring that phonetic notations are unique, variant romanization conventions in academic and community materials can be reconciled.

Romanization crosswalks. An additional possibility for consolidating variant romanization conventions could be to develop a "crosswalk" algorithm to programmatically navigate between multiple conventions. Using crosswalks, material in one convention can be converted into another based on specified rules. Once a standardized romanization convection is adopted by the community, a crosswalk program could be used to convert other conventions to the chosen standard. The practice of using crosswalks is commonly used in libraries to access and transfer data among repositories that use different metadata standards.

Ginan Dictionary. Surprisingly, an English dictionary of the ginans is yet to be developed and published. While there are several resources that feature back-of-the-book ginan glossaries with English meanings, the function of a dictionary is very different from a glossary. The dictionaries are typically developed for general audiences to provide grammatical information, definitions, synonyms, antonyms, as well as examples of usage. This is the reason why bilingual (or L2-to-L1) dictionaries are considered an indispensable resource for L2 learners (Laufer & Levitzky-Aviad, 2006).

The grammatical information that remains critical for L2 semantization is the word sense, which is a "discrete representation of one aspect of the meaning of a word" (Jurafsky & Martin, 2019, p. 355). The existing ginan glossaries typically list all the possible senses of a given word using single English words. For example, the following English synonyms are provided in the entry for the word āgal in the corpus used for this study: in front of; in the next world; ahead; leading; there. This list of words attempts to convey multiple senses of the word āgal. With a dictionary at their disposal, ginan learners would have access to "example sentences along with glosses, and these can again be used to help build a sense representation" (Jurafsky & Martin, 2019, p. 355).

If ginans are to be understood by English-speaking youth of the community, the availability of a ginan dictionary must become a priority for community elders and leaders. Developing a ginan dictionary can also be an opportunity to consolidate and standardize romanization for the ginans. Given the high preference in the community for online access, it may be worthwhile to make such a dictionary available online. In 2016, Wikimedia Foundation, the non-profit organization responsible for developing and hosting Wikipedia, unveiled an open

and free online dictionary platform called Wikibase Lexicographical data.[25] The platform may

be a useful starting point for the ginan dictionary project as it provides a multilingual data model

to manage lexicographic data.[26] In addition, the Wikibase Lexeme data model also includes

several attributes that have also been incorporated in the online lexicon developed in this

research, including parts of speech, gloss, word senses, etc.

---

[25] For more information and documentation on Wikidata Lexicographical data is available at
https://www.wikidata.org/wiki/Wikidata:Lexicographical_data
[26] The documentation on the Wikidata Lexemes and underlying data model is available at
https://www.wikidata.org/wiki/Wikidata:Lexicographical_data/Documentation

EPILOGUE: THE GIVING TRADITION

Useless are the tall trees that are without leaves,

So are the humans who are devoid of *ginan* (knowledge) from the *gur* (guide).

Translated from *Dhan dhan ājno,* a ginan attributed to Pir Sadardin

The religious community that Khojas Ismailis initially belonged to in India was called *Satpanthis* (followers of the True Path). The Satpanthis subscribe to the single spiritual reality of the human race which is agnostic to any particular religion, race, or practice. The *pirs* composed the ginans to spread this message using mixed language and drawing from the "bewildering thicket of Indian religions, mythologies and intellectual traditions… The ginans thus became and remained, until the contemporary project to reconceptualize and reformulate the Ismaili Tariqah (Tradition), the *de facto* supreme scripture for Satpanth Ismailis." (Alibhai, 2020, n.p.).

The colonization of India necessitated the reformulation of Satpanth as the "divide and rule" policy of the colonial rulers took hold of the Indian society (Asani, 2020, n.p.). The followers of Satpanth were forced to pick their religious affinity along the lines of two major religious persuasions – Hinduism and Islam. This caused turmoil among the followers of Satpanth as the reformulation of Satpanth in India dispersed into "a variety of identities along lines of 'monolithic' religious denomination: Nizari Ismailis, Ithna Ashari, Sunni, Hindus, devotees of Ramdev, etc." (Asani, 2020, n.p.).

The end of the colonial rule in 1947 resulted in the partition of the Indian subcontinent into two independent countries India and Pakistan. As the community's identity evolved from

Satpanthi Khojas to Nizari Ismaili Muslims, the "ginans [were] no longer the basis of normative understanding of being Nizari Ismaili; instead [ginans] became liturgical or ritualized texts" (Asani, 2020, n.p.). Consequently, the tradition of ginans has lost its status as the scripture of Satpanth and is now regarded as "devotional literature" of the Nizari Ismailis, at least from the institutional perspective.

This divergent and unaddressed evolution of perspectives on the ginans continue to exacerbate the anxiety about the future of gināns. Within the community, this division has resulted in debates on what constitutes official or authentic sources and materials of ginans. At the institutional level too, this dissonance has thwarted any sincere efforts to solicit and meet the devotional and intellectual needs of those community members who remain interested in the theological, ethical, and gnostic knowledge (*gyan*) encapsulated in the gināns.

Even as devotional literature, however, ginans continues to serve as a vehicle for community members to express their devotion through its poetry and melody. The emotive melodies (*rāgas*) of the *gināns* continue to bring faith, knowledge, devotion, and worship into ecstatic harmony even for those among the community who do not understand the language of the *gināns*. As Asani observes, "Even those who may not understand the meanings and significance of the words they sing may experience an emotion difficult to describe but which sometimes physically manifests itself through moist eyes or tears" (1992, p. 27).

One can draw a parallel between the fate of the tradition of ginans and that of the tree in Silverstein's parable titled *The Giving Tree*. This story features the life-long relationship between a generous tree and a selfish boy. As the boy grows from a little boy into a man, the generous tree is reduced to a stump as it selflessly gives away its fruits, branches, and finally its trunk to

110

make the boy happy through different stages of his life. Much like the boy in the story, the Khoja

Ismailis have drawn much more out of their relationship with the tradition of ginans throughout

their history than they have invested in nurturing the *giving tradition* of ginans.

REFERENCES

Alibhai, M. (2014). *Western India and Indian ocean studies: Long term vision.* Unpublished manuscript.

Alibhai, M. (2020, March 5). Tajbibi Abualy Aziz (1926-2019) Part one: A Satpanthi Sita. Retrieved from https://theolduvaireview.com/tajbibi-abualy-aziz/

Anthony, L. (2013). A critical look at software tools in corpus linguistics. *Linguistic Research*, 30(2), 141-161.

Asani, A. S. (2002). *Ecstasy and enlightenment: The Ismaili devotional literature of South Asia.* London: I.B. Tauris.

Asani, A. S. (2011). From Satpanthi to Ismaili Muslim: The articulation of Ismaili Khoja identity in South Asia. In F. Daftary (Ed.), *A modern history of the Ismailis: Continuity and change in a Muslim community* (pp. 95-128). London: I. B. Tauris Publishers.

Asani, A. S. (2020, December 20). *Towards a religiohistory of Ginans* [Webinar]. The Association for the Study of Ginans.

Asani, A. S. (2021). *The Ginans: Betwixt Satpanthī Scripture and "Ismaili" Devotional Literature* [Manuscript submitted for publication]. University Department, Harvard University.

Bakker, P. J. (2017). *Mixed languages.* Oxford Bibliographies. Retrieved February 21, 2021, from https://www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0030.xml

Bakker, P. J., & Mous, M. (1994). *Mixed languages. 15 case studies in language intertwining.* Amsterdam: IFOTT.

Barkwell, L. (2017). *A background paper on Michif: An overview of the last 35 years*.

    Unpublished manuscript. Downloaded from academic.edu. Retrieved February 20, 2021,

    from

    https://www.academia.edu/download/59510479/Background_on_Michif_and_Bibliograp

    hy_final_Barkwell20190604-9486-1cac9a5.pdf

Bateman, J.A., McDonald, D., Hiippala, T., Couto-Vale, D., & Costezki, E. (2017). Systemic-

    functional linguistics and computation: new directions, new challenges. In G. Thompson,

    W. L. Bowcher & L. Fontaine, J. Y. Liang (Eds.), *The Cambridge handbook of systemic*

    functional linguistics. Cambridge: Cambridge University Press.

Beheydt, L. (1987). The semantization of vocabulary in foreign language learning. *System*, *15*(1),

    55-67.

Berners-Lee, T. (2012). *History of the web*. The World Wide Web Foundation. Retrieved

    February 16, 2019, from https://webfoundation.org/about/vision/history-of-the-web

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific*

    American, *284*(5), 34-43.

Borg, W., & Gall, M. (2003). *Educational research: An introduction* (5th ed.). New York:

    Longman.

Bowker, L. (2018). Corpus linguistics is not just for linguists: Considering the potential of

    computer-based corpus methods for library and information science research. *Library Hi*

    Tech, *36*(2), 358-371.

Bowker, L., & Pearson, J. (2002). *Working with specialized language: a practical guide to using*

    corpora. Routledge.

Breyer, Y. (2009). Learning and teaching with corpora: reflections by student teachers. *Computer assisted language learning, 22(2),* 153-172.

Burger A., & Chong I. (2011) Receptive vocabulary. In S. Goldstein & Naglieri J.A. (Eds.), *Encyclopedia of child behavior and development* (pp.1231-1231). Springer, Boston, MA.

Busa, R. (1980). The annals of humanities computing: The index Thomisticus. *Computers and the Humanities*, *14*(2), 83-90.

Chomsky, N. (2006). *Language and mind*. Cambridge: Cambridge University Press.

Collins Online English Dictionary, (2020). Concordance. In *Collins Online English Dictionary*. Retrieved December 13, 2020, from

https://www.collinsdictionary.com/dictionary/english/concordance

Collins Online English Dictionary. (2020). Diacritic. In *Collins online English dictionary*. Retrieved February 18, 2021, from

https://www.collinsdictionary.com/dictionary/english/diacritic

Collins Online English Dictionary. (2020). Diaspora. In *Collins online English dictionary*. Retrieved July 5, 2021, from

https://www.collinsdictionary.com/dictionary/english/diaspora

Collins Online English Dictionary. (2020). Lexicon. In *Collins online English dictionary*. Retrieved July 9, 2021, from

https://www.collinsdictionary.com/dictionary/english/lexicon

Collins Online English Dictionary. (2020). Polyglottic. In *Collins online English dictionary*. Retrieved December 13, 2020, from

https://www.collinsdictionary.com/dictionary/english/polyglottic

Collins Online English Dictionary. (2020). Proof-of-concept. In *Collins online English dictionary*. Retrieved December 13, 2020, from

https://www.collinsdictionary.com/dictionary/english/proof-of-concept

Collins Online English Dictionary. (2020). Wireframe. In *Collins online English dictionary*. Retrieved December 13, 2020, from

https://www.collinsdictionary.com/dictionary/english/wireframe

Cronbach, L. (1942). An analysis of techniques for diagnostic vocabulary testing. *The Journal of Educational Research, 36*(3), 206-217.

Daftary, F. (2011). *Historical dictionary of the Ismailis*. New York: Scarecrow Press.

Dale, E. (1965). Vocabulary measurement: techniques and major findings. *Elementary English, 42,* 82-88.

Daller, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.

Fitzpatrick, K. (2012). The humanities, done digitally, debates in the digital humanities. In M.K. Gold. (Eds.), *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota Press.

Flanders, J. (2013). The productive unease of 21st-century digital scholarship. *DHQ: Digital Humanities Quaterly, 3*(3), 205-218.

Freitas, A., Curry, E., Oliveira, J. G., & O'Riain, S. (2012). Querying heterogeneous datasets on the linked data web: challenges, approaches, and trends. *IEEE Internet Computing*, *16*(1), 24-33.

Furuta, R., Shipman III, F. M., Marshall, C. C., Brenner, D., & Hsieh, H. W. (1997, April). Hypertext paths and the World-Wide Web: experiences with Walden's paths. In *Proceedings of the eighth ACM conference on Hypertext*, 167-176. ACM.

Goddard, C. (2011). *Semantic analysis: A practical introduction*. Oxford: Oxford University Press.

Harispe, S., Ranwez, S., Janaqi, S., & Montmain, J. (2015). Semantic similarity from natural language and ontology analysis. *Synthesis Lectures on Human Language Technologies*, *8*(1), 1-254. Retrieved November 22, 2018, from https://arxiv.org/pdf/1803.03198.pdf

Henriksen, B. (1999). Three dimensions of vocabulary development. *SSLA, 21,* 303-317.

Holmes, W. (2013). *Level Up! A design-based investigation of a prototype digital game for children who are low-attaining in mathematics*. Doctoral thesis (D.Phil.) University of Oxford. Retrieved from https://solo.bodleian.ox.ac.uk.

Horrocks, I., Kagal, L., & Hotho, A. (2017). Web semantics for digital humanities. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web, 44,* 1-118.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Internet Live Stats. (2017). *Real time statistics project*. Retrieved February 16, 2018, from https://www.internetlivestats.com/total-number-of-websites

Izutsu, T. (1964). *God and man in the Koran: Semantics of the Koranic Weltanschuung*. Japan: Keio Institute of Cultural and Linguistics Studies.

Izutsu, T. (2002). *Ethico-religious concepts in the Qur'an*. Montreal: McGill University Press.

Jewitt, C., & Kress, G. R. (Ed.). (2003). *Multimodal literacy*. New York: P. Lang.

Jurafsky, D., & Martin, J. H. (2019). *Speech & language processing*. Unpublished manuscript.

    Draft of October 2, 2019. Retrieved from

    https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Karim, K. H. (2021, January 10). *Institute of Ismaili Studies: Historical aspirations,*

    contemporary possibilities. Simerg – Insights from Around the World.

    https://simerg.com/2021/01/10/institute-of-ismaili-studies-historical-aspirations-

    contemporary-possibilities/

Klein, J. T., & Newell, W. H. (1996). Advancing interdisciplinary studies. In J. Gaff & J.

    Ratcliffe (Eds.), *Handbook of the undergraduate curriculum: a comprehensive guide to*

    purposes, structures, practices, and change (pp. 393-415). Jossey-Bass, San Francisco.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.).

    Thousand Oaks, CA: Sage.

Laufer, B., & Levitzky-Aviad, T. (2006). Examining the effectiveness of 'bilingual dictionary

    plus'–a dictionary for production in a foreign language. *International Journal of*

    Lexicography, *19*(2), 135-155.

Leech, G. (1997). Introducing corpus annotation. In R. Garside, G. Leech, A. McEnry (Eds),

    Corpus Annotation, pp. 1-18. London: Longman.

Lenders, O. (2008). Electronic glossing: Is it worth the effort? *Computer Assisted Language*

    Learning, 21(5), 457-481.

Lewis, M. P., Simons, G. F., & Fennig, C. D. (2020). *Mixed languages*. Ethnologue: languages

    of the world. Retrieved February 21, 2021, from

    https://www.ethnologue.com/subgroups/mixed-language

Libben, G. (1990). Noam Chomsky: Consensus and controversy by Sohan Modgil and Celia

      Modgil. *Canadian Journal of Education / Revue Canadienne De L'éducation*, *15*(4), 462-

      464.

Marshall, C. C., & Shipman, F. M. (2003). Which semantic web? In *Proceedings of the*

      fourteenth ACM conference on Hypertext and hypermedia, 57-66. ACM.

Martinez, A. (2010). Natural language processing. *Wiley Interdisciplinary Reviews:*

      Computational Statistics, 2(3), 352-357.

McCarty, W. (2015). The Big Picture: Where digital humanities has been & where you might

      take it. Workshop for the *Digital Humanities Conference*, Siberian Federal University,

      Krasnoyarsk. Retrieved November 25, 2018, from https://www.mccarty.org.uk

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced*

      resource book. London, UK: Routledge, Taylor & Francis Group.

Melgosa, A. A. (2018). *Developing a faculty-librarian community of inquiry: A blended learning*

      approach to facilitate information literacy education [Doctoral dissertation, University of

      Calgary].

Minsky, M. (1967). *Computation: Finite and infinite machines*. Englewood Cliffs, NJ: Prentice-

      Hall.

Mishra, S. K. (Ed.). (2018). *Artificial intelligence and natural language processing*. Cambridge,

      UK: Cambridge Scholars Publishing.

Mixed language. (2020, November 19). Mixed language. In *Wikipedia*.

      https://en.wikipedia.org/w/index.php?title=Mixed_language&oldid=989600726

ßMoretti, F. (2013). *Distant reading*. London: Verso Books.

Muñoz, T. (2013). Data curation as publishing for the digital humanities. *Journal of Digital Humanities*, *2*(3), 14-22.

Murdock, J., Allen, C., Börner, K., Light, R., McAlister, S., Ravenscroft, A., et al. (2017). Multi-level computational methods for interdisciplinary research in the HathiTrust Digital Library. *PloS one*, *12*(9), e0184188.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada journal*, 09-29.

Petten, C. (2006). Michif speakers talk language preservation. *Ontario Birchbark*, *5*(4), 1772-1781. Retrieved February 20, 2021, from https://ammsa.com/publications/ontario-birchbark/michif-speakers-talk-language-preservation

Poole, A. H. (2017). A greatly unexplored area: Digital curation and innovation in digital humanities. *Journal of the Association for Information Science and Technology*, *68*(7), 1772-1781.

Poole, R. E. (2011). *Concordance-based glosses for facilitating semantization and enhancing productive knowledge of academic vocabulary* [Doctoral dissertation, University of Alabama Libraries].

Posner, M. (2018). Digital Humanities. In M. Kackman & Kearney, M. C. (Eds.), *The craft of criticism: Critical media studies in practice*. New York: Routledge.

Pownell, D. W. (2002). *Implementing handheld computers in schools: The research, development, and validation of a technology leader's resource guide* [Doctoral dissertation, Kansas State University].

Romanization. (2020, December 12). Romanization. In *Wikipedia.* https://en.wikipedia.org/w/index.php?title=Romanization&oldid=993742447

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329-363.

Schreier, D. (2015). Review of the book *Contact languages: A comprehensive guide*, by In P. Bakker & Y. Matras. *World Englishes*, *35*(3), 480-483.

Seadle, M. S. (2016). Managing and mining historical research data. *Library Hi Tech*, *34*(1), 172-179.

Shackle, C., & Moir, Z. (2000). *Ismaili hymns from South Asia: An introduction to the Ginans.* Richmond, Surrey: Curzon Press.

Shadbolt, N., Berners-Lee, T., & Hall, W. (2006). The semantic web revisited. *IEEE intelligent systems*, *21*(3), 96-101.

Sharma, H., & Sharma, A. K. (2017). Study and analysis of topic modelling methods and tools– A survey. *American Journal of Mathematical and Computer Modelling*, *2*(3), 84-87.

Simon, H. (1965). *The shape of automation for men and management*, New York: Harper & Row.

Tai, K. S., Socher, R., & Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint*

Conference on Natural Language Processing (Volume 1: Long Papers), pp. 1556–1566, Beijing, China. Association for Computational Linguistics.

Tharani, K. (2019). Shifting established mindsets and praxis in libraries: Five insights for making non-Western knowledge digitally accessible through community engagement. *Canadian Journal of Academic Libraries*, *4*(February), 1-13.

The Ismaili Muslim Community. (2020). The official website of the Ismaili Muslim community. Retrieved December 13, 2020, from https://the.ismaili/

Transcription. (2019, July 17). In *Wikipedia.*

https://en.wikipedia.org/w/index.php?title=Transcription&oldid=906668106

Transliteration. (2020, November 30). In *Wikipedia.*

https://en.wikipedia.org/w/index.php?title=Transliteration&oldid=991544704

Unsworth, J. (2002). What is humanities computing and what is not*?* In M. Terras, Nyhan, J., & Vanhoutte, E. (Eds.), *Defining Digital Humanties – A Reader.* Surrey: Ashgate.

Van den Akker, J. (1999). Principles and methods of development research. In J. Van den Akker, Branch, R. M., Gustafson, K., Nieveen, N., & Plomp, T. (Eds.), *Design approaches and tools in education and training* (pp. 1-14). Dordrecht: Springer.

Virani, S. (2015). Introduction. In *Ginans with English translation and glossary – Volume 9* (pp. viii-xvi). Pakistan.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development*, *53*(4), 5-23.

Waring, R. (2002). Scales of vocabulary knowledge in second language vocabulary assessment. Appeared in *Kiyo, The Occasional Papers of Notre Dame Seishin University*, March 2002.

Weaver, W. (1949). The mathematics of communication. *Scientific American*, *181*(1), 11-15.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, *30,* 79-95.

Yang, M. C., & Rau, D. V. (2005). An integrated framework for archiving, processing and developing learning materials for an endangered aboriginal language in Taiwan. In *Proceedings of the Fifth Workshop on Asian Language Resources (ALR-05) and First Symposium on Asian Language Resources Network (ALRN)*.

APPENDIX A: RESEARCH CORPUS AND GINANS

Research Corpus

The following resources were used as the research corpus for this study:

1. Alwaez Kamaluddin, Muhammad. (2011). *50 Ginans with English Translation and Glossary - Volume 1* (Second edition).

2. Alwaez Kamaluddin, Muhammad. (1990). *Concordance of the Ginans - Volume 1*.

3. Princeton University. (2010). *Online WordNet. WordNet*. Princeton University. Accessible at https://wordnetweb.princeton.edu/perl/webwn

Sample Ginans

The following ten ginans were selected for this study:

1. sāchā merā khālak (6 verses)

2. kāyam dāyam tuṅ moro sāmī (6 verses)

3. uṅchathī āyo baṅde nīch kyuṅ dhīāve (4 verses)

4. ab terī mahobat lāgī (10 verses)

5. sabhāgā is dhunīāṅdhe vich kiyā gin āve (11 verses)

6. sabhāgā har puj nit nīravāṇ paṅjetan (11 verses)

7. kesarī siṅh svaroop bhulāyo (5 verses)

8. suno suno bhāīve momano (16 verses)

9. ek tīrath vedhaṇā pīr shams gazī sadhaṇā (6 verses)

10. bhulo bhulo te bhulo bhamarḍore lol (22 verses)

APPENDIX B: ONLINE SURVEY CONSENT AND QUESTIONNAIRE

Participant Consent Form

Researcher: Karim Tharani
Associate Librarian,
University Library
University of Saskatchewan
Room 163, 3 Campus Drive, Saskatoon SK S7N 5A5
Phone: 306-966-8785
Email: karim.tharani@usask.ca

Research Supervisor: Dr. Jay Wilson
Associate Professor & Department Head,
Department of Curriculum Studies, College of Education
University of Saskatchewan
Room 3023, 28 Campus Drive, Saskatoon, SK S7N 0X1
Phone: 306-966-7617
Email: jay.wilson@usask.ca

Procedure: You will be asked to share your needs and preferences in an online survey. The estimated time commitment to fill out the survey is between 10 to 15 minutes.

Risks and Discomforts: We do not foresee any risks or discomfort from your participation in the research. You have the right to not answer any questions.

Benefits of the Research and Benefits to You: We believe your experiences and insights as a learner of ginans are highly valuable as we look to combine digital technologies and library collections to better meet the needs of community members interested in understanding ginans. With the information you provide, we will develop online learning resources to make ginans more accessible and comprehensible.

Confidentiality: Data collected from this survey will be reported in aggregated form only, potentially with some anonymized direct quotations. No personal identifying information will be shared and the participant's identity will not be known to anyone. However, the researchers will have access to individual responses and may be able to identify responses from individuals where contact information has been provided.

Storage of Data: This survey data is hosted by SurveyMonkey using the University of Saskatchewan's enterprise account, which ensures that data is securely stored in Canada. Please see Survey Monkey's Privacy Policy for more details. Data will be stored for a minimum of five years post-report, or post-publication, whichever is longer.

Right to Withdraw: Participation in this survey is voluntary. You can decide not to participate at any time by closing your browser or choose not to answer any questions you do not feel comfortable with.

Follow up: We may contact you for this or other research on ginans, provided you have filled in your

contact information details in the survey. Final reports and or publications may be shared with the University of Saskatchewan community via emails, websites, and social media. Final reports and/or publications may also be disseminated in academia.

Questions or Concerns: Contact the researchers using the information at the top of this page. This research project has been approved on ethical grounds by the University of Saskatchewan Behavioural Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office: ethics.office@usask.ca; 306-966-2975; out of town participants may call toll free 1-888-966-2975.

By completing and submitting this questionnaire, your free and informed consent is implied and indicates that you understand the above conditions of participation in this study.

Please consider printing this page for your records.

Survey Questionnaire

1. In what country do you live?

[   ]

Other (please specify)

[   ]

2. Which of the following age group do you belong to?

○ 18 to 24 years old

○ 25 to 34 years old

○ 35 to 44 years old

○ 45 to 54 years old

○ 55 to 64 years old

○ 65 to 74 years old

○ Over 75 years old

3. What is your gender?

○ Female

○ Male

○ Other

4. What is the <u>highest</u> level of school or degree you have received?

○ No formal education

○ Less than high school diploma

○ High school diploma or equivalent

○ Vocational training

○ Some college but no degree

○ College degree

○ Bachelor's degree

○ Master's degree

○ Doctorate degree or higher

○ Other (please specify below)

[   ]

5. What is your primary occupation?

( ) Business owner

( ) Employed

( ) Homemaker

( ) Retired

( ) Self-employed

( ) Student

( ) Unemployed

( ) Other (please specify below)

[   ]

6. Are you a member of the Ismaili community?

( ) Yes

( ) No

7. How important are ginans in your life?

Not at all	A little	A moderate amount	A lot	A great deal
( )	( )	( )	( )	( )

8. Are you concerned that the tradition of ginans may be lost in the future?

Not at all	A little	A moderate amount	A lot	A great deal
( )	( )	( )	( )	( )

Why or why not?

[   ]

9. Are you concerned that the teachings of ginans may be lost in the future?

Not at all	A little	A moderate amount	A lot	A great deal
( )	( )	( )	( )	( )

Why or why not?

[   ]

10. Do you speak most often in English at home?

( ) Yes

( ) No

11. Which languages are you able to read or write in?

- ☐ Khojki
- ☐ Gujarati
- ☐ English
- ☐ Urdu
- ☐ Other (please specify)

[_____]

12. What type of content do you currently use for ginans?

	Not at all	A little	A moderate amount	A lot	A great deal
Ginan texts	○	○	○	○	○
Ginan translations	○	○	○	○	○
Ginan audios	○	○	○	○	○
Ginan videos	○	○	○	○	○
Academic articles & presentations	○	○	○	○	○
Community talks & sermons (waez)	○	○	○	○	○
Glossaries & dictionaries	○	○	○	○	○
Other	○	○	○	○	○

Other (please specify)

[_____]

13. How well do you understand the ginans?

Not at all	A little	A moderate amount	A lot	A great deal
○	○	○	○	○

14. How interested are you in understanding the ginans?

Not at all	A little	A moderate amount	A lot	A great deal
○	○	○	○	○

15. How important is it for you to be able to access content online to learn and understand ginans?

Not at all	A little	A moderate amount	A lot	A great deal
○	○	○	○	○

128

16. Which of the following sources do you use for accessing ginan materials <u>online</u>?

	Not al all	A little	A moderate amount	A lot	A great deal
Ginan Central (University of Saskatchewan Library)	○	○	○	○	○
The Association for the Study of Ginans (GINANS.org)	○	○	○	○	○
The Ismaili Literature in Indic Languages (Harvard University)	○	○	○	○	○
The Heritage Society (Ismaili.NET)	○	○	○	○	○
The Institute of Ismaili Studies (IIS)	○	○	○	○	○
Independent websites, blogs, YouTube, social media, etc.	○	○	○	○	○
Other	○	○	○	○	○

Other (please specify)

[   ]

17. What language/script do you prefer to use for <u>learning and understanding</u> the ginans?

	Not at all	A little	A moderate amount	A lot	A great deal
Khojki	○	○	○	○	○
Gujarati	○	○	○	○	○
English	○	○	○	○	○
Urdu	○	○	○	○	○
Other	○	○	○	○	○

Other (please specify)

[   ]

129

18. Which of the following online content will be useful for you to learn and understand the ginans?

	Not at all	A little	A moderate amount	A lot	A great deal
Ginan texts with translations	○	○	○	○	○
Ginan texts with links to glossary	○	○	○	○	○
Ginan audios with tune (recited)	○	○	○	○	○
Ginan audios without tune (spoken)	○	○	○	○	○
Ginan videos with text, translation, audio, etc.	○	○	○	○	○
Ginan gist/overview	○	○	○	○	○
Ginan dictionary	○	○	○	○	○
Ginan categories	○	○	○	○	○
Other (please specify below)	○	○	○	○	○

Other (please specify)

[   ]

19. What kind of device would you most prefer to use to access ginan materials online?

	Not at all	A little	A moderate amount	A lot	A great deal
Computer tablet	○	○	○	○	○
Desktop computer	○	○	○	○	○
Smart phone	○	○	○	○	○
Other	○	○	○	○	○

Other (please specify)

[   ]

20. In your opinion what additional measures can be taken to ensure that the tradition and teachings of ginans remain available to future generations?

[   ]

Thank you again for taking the time to complete this survey!

21. Please provide us with your contact information if you would like us to contact you for future research on ginans.

Name

Email Address

APPENDIX C: INVITATION AND CONSENT FOR COMMUNITY EDUCATORS

<u>Sample Email Invitation</u>

Subject: Invitation to Participate – Online Survey, Ginan Research @ U of S

Dear XXX,

This email is to invite you to participate in an online survey to help evaluate an online tool for learning and understanding ginans. The purpose of this survey is to collect feedback from community members like you who will be the end-users of the tool.

We believe your experiences and insights as a learner of ginans will be highly valuable as we look to incorporate digital technologies and library collections and archives to better meet the needs of the Ismaili community in engaging with and learning the tradition of ginans. With the information you provide, we will refine the tool to make it as effective as possible for the wider Ismaili community members.

The online survey will only take 10-20 minutes and will available starting on <date>. If you are willing to participate in the online survey, please confirm you willingness to participate via return email to Karim Tharani (karim.tharani@usask.ca). You will be asked to provide consent at the beginning of the online survey.

If you have any questions or concerns about the research, your participation, or your consent, please do not hesitate to contact me directly via email or phone at 306-966-8785. We will also have an opportunity to go over the consent form at the beginning of the focus group session. Please note that you can withdraw your consent at any time by letting me know. We will also take all appropriate measures to maintain confidentiality of all the information you provide us, which will only be used for the purposes of this research.

Your contribution is key and look forward to hearing from you!

Sincerely,
Karim Tharani
Doctoral Candidate,
Department of Curriculum Studies - College of Education
Room 163, 3 Campus Drive, Saskatoon SK S7N 5A5
Email: karim.tharani@usask.ca Tel: 306-966-8785

<u>Consent Form</u>

Study Name: At the Nexus of Tradition and Technology: A design-based investigation of a prototype semantization tool for mixed-language of ginans

Researcher: Karim Tharani is the researcher for this study. His contact information is as follows:

> Karim Tharani
> Doctoral Candidate,
> Department of Curriculum Studies, College of Education
> University of Saskatchewan
> Room 163, 3 Campus Drive, Saskatoon SK S7N 5A5
> Phone: 306-966-8785
> Email: karim.tharani@usask.ca

Research Supervisor: Dr. Jay Wilson is the research supervisor for this research. His contact information is as follows:

> Dr. Jay Wilson
> Professor & Department Head,
> Department of Curriculum Studies, College of Education
> University of Saskatchewan
> Room 3023, 28 Campus Drive, Saskatoon, SK S7N 0X1
> Phone: 306-966-7617
> Email: jay.wilson@usask.ca

Purpose of the Research: As part of my PhD research I will be designing an online tool based on the existing semantization theories and best practices. I will be working with members of the Ismaili community and their collections of gnostic and religious hymns known as "ginans". The language of ginans is both oral (with no specific script) and mixed (with vocabulary from Arabic, Persian, and multitude Indic languages). The purpose of my research is to utilize semantization best practices and technology to digitally curate available learning materials in the form of a prototype of an online self-learning tool for community learners to learn and understand ginan vocabulary.

What You Will Be Asked to Do in the Research: As a participant, you will be asked to participate in an online interview to provide your thoughts on the best practices and materials used for teaching ginans. During the interview you will also have the opportunity to highlight the current challenges that instructors and learners face when engaging with the ginans. There will be video recording of the interview which will be used for transcribing our discussion for research purposes. The recordings will be handled with strict confidentiality and will only be available to those closely related to this research.

Your interview will be on **<time, date>** at **<venue>**. The online will last approximately 60-90 minutes. In order to participate, you must

1. **have been or are directly involved in transmission and/or teaching of ginans, and,**
2. **must have direct knowledge of the issues and challenges faced by teachers and learners**

Risks and Discomforts: We do not foresee any risks or discomfort from your participation in the research. You have the right to not answer any questions.

Benefits of the Research and Benefits to You: We believe your experiences and insights as an instructor of ginans are highly valuable as we look to incorporate digital technologies and library collections and archives to better meet the needs of community youth in engaging with the tradition of ginans. With information you provide, we will develop and fine tune the tool to make it as effective as possible for the wider community members.

Voluntary Participation: Your participation in the study is completely voluntary and you may choose to stop participating at any time. Your decision not to participate, or withdraw your consent at any time, will not influence the nature of our relationship either now, or in the future.

Withdrawal from the Study: You can stop participating in the study at any time, for any reason, if you so decide. Your decision to stop participating or to withdraw our data, will not affect your relationship with the researcher or anyone associated with this project. While advance notice is always appreciated, your right to withdraw data from the study will apply until within one month of your interview date. After this, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data. In order to withdraw, please contact Karim Tharani either by telephone at (306) 966-8785 or by e-mail (karim.tharani@usask.ca).

Confidentiality: You interview will be conducted by videoconference and will be auto-recorded for transcription. You may request that the auto-recorder be turned off at any time. The digital recording of your interview will remain confidential and will not be shared publicly. The data from your interview will be transcribed and will be disseminated in aggregate and summarized form via thesis, articles, and/or report to an agency or community. Any direct quotes will be anonymized to ensure participant's confidentiality. The confidential digital recordings will be treated as confidential data and will be guided by the research data storage and retention policy.

Data Storage and Retention: Your data will be safely stored in a locked facility at the University of Saskatchewan and only research staff will have access to this information. The data will be stored for five years <u>post publication</u> and will be destroyed thereafter.

Research Results: The results of this research will be disseminated via thesis, articles, and/or report to an agency or community. Participants can also request information on the research results by contacting Karim Tharani either by telephone at (306) 966-8785 or by e-mail ([karim.tharani@usask.ca)](karim.tharani@usask.ca).

Questions About the Research? If you have questions about the research in general or about your role in the study, please feel free to contact Karim Tharani either by telephone at (306) 966-8785 or by e-mail ([karim.tharani@usask.ca)](karim.tharani@usask.ca). This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

Oral Consent:
I read and explained this consent form to the participant before receiving the participant's consent, and the participant had knowledge of its contents and appeared to understand it.

_____   _____   _____
   Name of Participant   **Researcher's Signature**   **Date**

APPENDIX D: LIST OF GUIDING QUESTIONS FOR EXPERT REVIEWS

Review questions:

(1) What resources do you typically rely on when learning ginans?

(2) What challenges do you generally feel when learning ginans in person from a teacher?

(3) Would the proposed ginan vocabulary semantization tool prove to be useful for you?

(4) What is the greatest strength of the tool for learning ginan vocabulary?

(5) What is the greatest weakness of the tool for learning ginan vocabulary?

(6) What features would you add or remove to improve the tool for vocabulary semantization?

(7) What additional learning resources would you like to be added to the tool?

(8) Are the content, interface, and navigational feature conducive to online semantization of the ginan vocabulary?

(9) What other comments or suggestions would you like to make about the tool?

APPENDIX E: VALIDATION SURVEY CONSENT AND QUESTIONNAIRE

Online Ginan Learning Tool - Validation Survey

Welcome!

Thank you for participating in Ginan Survey 2020 and volunteering to be contacted for our research on Ginans at the University of Saskatchewan.

This time we are inviting you to help us evaluate our online Ginan learning tool. This tool has been designed based on the needs of the community gathered through Ginan Survey 2020 and with input from Ginan instructors and experts. The tool also brings together the best practices for online learning adapted for the oral tradition of Ginan.

In order to participate in this study, we will need a few minutes of your time to do the following:

Step 1
On the next page, you will be presented with a form to provide us with your consent to collect information from you for this study.

Step 2
After providing your consent, you will be given a link to access the online Ginan learning tool. Please spend some time exploring the tool to assess its usefulness as a tool for you to learn and understand Ginans online.

Step 3
After spending some time with the tool, you can return here to fill out a short online survey to provide us with your feedback on the online Ginan learning tool.

Step 1: Providing Consent

Researcher: Karim Tharani
University Library
University of Saskatchewan
Room 163, 3 Campus Drive, Saskatoon SK S7N 5A5
Phone: 306-966-8785
Email: karim.tharani@usask.ca

Research Supervisor: Dr. Jay Wilson
Professor & Department Head,
Department of Curriculum Studies, College of Education
University of Saskatchewan
Room 3023, 28 Campus Drive, Saskatoon, SK S7N 0X1
Phone: 306-966-7617
Email: jay.wilson@usask.ca

Risks and Discomforts: We do not foresee any risks or discomfort from your participation in the research. You have the right to not answer any questions.

Benefits of the Research and Benefits to You: We believe your experiences and insights as a learner of Ginans are highly valuable as we look to combine digital technologies and library collections to better meet the needs of community members interested in understanding Ginans. With the information you provide, we will develop resources (materials, tools, and technology) to make Ginans and their meanings more accessible to the wider community members interested in learning Ginans.

Confidentiality: Data collected from this survey will be reported in aggregated form only, potentially with some direct quotations; however, no personal identifying information will be shared and the participant's identity will not be known to anyone. The researchers will have access to individual responses and may be able to identify responses from individuals with the contact information provided.

Storage of Data: This survey data is hosted by SurveyMonkey using the University of Saskatchewan's enterprise account, which ensures that data is securely stored in Canada. Please see SurveyMonkey's Privacy Policy for more details. Data will be stored for a minimum of five years post-report, or post-publication, whichever is longer.

Right to Withdraw: Participation in this survey is voluntary. You can decide not to participate at any time by closing your browser or choose not to answer any questions you do not feel comfortable with.

Questions or Concerns: Contact the researchers using the information at the top of this page. This research project has been approved on ethical grounds by the University of Saskatchewan Behavioural Research Ethics Board. Any questions regarding your rights as a participant may be

addressed to that committee through the Research Ethics Office: ethics.office@usask.ca; 306-966-2975; out of town participants may call toll free 1-888-966-2975.

By completing and submitting this questionnaire, your free and informed consent is implied and indicates that you understand the above conditions of participation in this study.

We strongly recommend that you print this page and retain a copy of it for your records.

Step 2: Accessing and Exploring the Online Ginan Tool

Accessing the tool

You can access the online Ginan learning tool by clicking on the link provided below. Please do not close this screen as you will need to return to this screen once you are ready to provide your feedback on the tool. If you close this screen accidentally, you can use the same link from your invitation email to return to this screen. Clicking the following link will open the tool will open in a separate window: Online Ginan Learning Tool

Exploring the tool

The purpose of this tool is to provide all the necessary resources for learners to be able to learn and understand Ginans online. Please spend as much time as you need to explore the tool to give us your feedback on the usefulness of this tool to you. You can use the following strategies to explore the tool and its features.

- Explore Ginans that are most interesting, and perhaps even unfamiliar, to you from the available list of sample Ginans.
- Click on the "Back" button to return to the Ginan list.
- Explore various options in the toolbar by clicking each of the buttons.
- Listen to the recital of your chosen ginan by clicking the play button of the audio player.
- Access the pronunciation of special letters (such as ā, ī, ū, etc.) by clicking on the Pronunciation Guide link.
- Review the side-by-side and verse-by-verse text and translation of the Ginans.
- Read the English text of the ginans and hover over and click on the linked words to access more information.
- Explore the semantic profiles of various linked Ginanic terms.
- Learn to pronounce difficult words by clicking on the speaker icon in the semantic profile of the words.
- Read the English meanings and synonyms of important terms in the ginan presented in the glossary.

Online Ginan Learning Tool - Validation Survey

Step 3: Survey Questionnaire

Please provide us with your feedback by answering the following questions about the online Ginan tool.

1. What kind of device did you use to access the online tool?

Smart phone	Desktop computer	Laptop computer	Tablet or pad	Other
○	○	○	○	○

2. Approximately how much time did you spend exploring the online tool?

None	Less than 5 minutes	Between 5 and 15 minutes	Between 15 and 30 minutes	Over 30 minutes
○	○	○	○	○

3. The online tool provides access to resources that I may need to learn and understand Ginans effectively.

Strongly disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Strongly agree
○	○	○	○	○

4. The overall layout and interface of the online tool are easy to use and follow.

Strongly disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Strongly agree
○	○	○	○	○

5. All the links, icons, and navigation buttons worked as expected on my device.

Strongly disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Strongly agree
○	○	○	○	○

6. I will use the online Ginan tool when it is made available to the community.

Strongly disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Strongly agree
○	○	○	○	○

7. What other comments or suggestions would you like to make about the online tool?

Online Ginan Learning Tool - Validation Survey

Thank you for taking the time to complete this survey!

We deeply appreciate your time and thoughtful feedback.