|  |
| --- |
| [Title]  Wilbu Wildcat  Department of Linguistics, The University of Arizona  Course Number: Course Title  Instructor Title and Name  [Publish Date] |
| Notice Regarding the Use of AI Technologies in Assignment Preparation  This assignment may incorporate contributions from artificial intelligence (AI) technologies, such as [ChatGPT Plus](https://openai.com/blog/chatgpt-plus/), [Microsoft Copilot Pro](https://www.microsoft.com/en-us/store/b/copilotpro), [Grammarly Premium](https://www.grammarly.com/premium), and [GitHub Copilot](https://github.com/features/copilot/). Significant AI-generated content, where used, has been cited following [APA](https://apastyle.apa.org/blog/how-to-cite-chatgpt) or similar standards, including links to original prompts when feasible, as in the citation for this paragraph. The roles of these AI tools in research and composition are documented, underscoring a commitment to transparency and academic integrity. Responsibility for the content's accuracy and quality rests with the student, acknowledging that AI-generated material has been vetted for inaccuracies and biases. This notice serves as an acknowledgment of AI assistance, aligning with our university’s academic standards and fostering a culture of integrity in using advanced technologies as published at <https://artificialintelligence.arizona.edu/> ([ChatGPT-4, 2024](https://chat.openai.com/share/56182658-3765-48f4-ba33-0569ae0cc4ab)). |

Abstract

{One-paragraph summary of the entire paper – typically no more than 250 words in length (and usually well shorter than that), the Abstract provides an overview of the study. The abstract should stand alone and contain no information that is not in the main body of the paper. Abstracts are optional for student papers. Check assignment instructions. Delete this entire page if an abstract is not required.}

Table of Contents

[☐ Background on CNL and CTI {August 2024} 6](#_Toc174081343)

[☐ Introduction to CNLs 6](#_Toc174081344)

[☐ Introduction to CTI and STIX 7](#_Toc174081345)

[☐ Intersection of CNL and CTI 9](#_Toc174081346)

[☐ Objectives and Scope {July 2024} 11](#_Toc174081347)

[☐ Purpose: 11](#_Toc174081348)

[☐ Users and Requirements 11](#_Toc174081349)

[☐ Tasks 11](#_Toc174081350)

[☐ Scope 11](#_Toc174081351)

[☐ Existing Languages and Standards {September 2024} 12](#_Toc174081352)

[☐ Existing STIX-Ds 12](#_Toc174081353)

[☐ Domain Standards 12](#_Toc174081354)

[☐ Best Practices 12](#_Toc174081355)

[☐ Corpus {September 2024} 13](#_Toc174081356)

[☐ Source Documents 13](#_Toc174081357)

[☐ Annotation and Analysis 13](#_Toc174081358)

[☐ Corpus Management 13](#_Toc174081359)

[☐ Lexicon {August 2024} 14](#_Toc174081360)

[☐ Key Concepts and Terms 14](#_Toc174081361)

[☐ Definitions 14](#_Toc174081362)

[☐ Consistency 14](#_Toc174081363)

[☐ Grammar Rules {Q4, 2024} 15](#_Toc174081364)

[☐ Syntax 15](#_Toc174081365)

[☐ Semantics 15](#_Toc174081366)

[☐ Testing 15](#_Toc174081367)

[☐ Semantic Framework {Q1, 2025} 16](#_Toc174081368)

[☐ {ADD SECTION HEADING} 16](#_Toc174081369)

[☐ {ADD SECTION HEADING} 16](#_Toc174081370)

[☐ Testing 16](#_Toc174081371)

[☐ Tools and Resources {Q2, 2025} 17](#_Toc174081372)

[☐ Authoring Tools 17](#_Toc174081373)

[☐ Parsing Tools 17](#_Toc174081374)

[☐ Validators 17](#_Toc174081375)

[☐ Testing and Evaluation {Q2, 2025} 18](#_Toc174081376)

[☐ {Add Heading} 18](#_Toc174081377)

[☐ {Add Heading} 18](#_Toc174081378)

[Improve 18](#_Toc174081379)

[☐ Documentation and Training {Q3, 2025} 19](#_Toc174081380)

[☐ Documentation 19](#_Toc174081381)

[☐ Training Materials 19](#_Toc174081382)

[☐ {Add Heading} 19](#_Toc174081383)

[☐ Deployment and Maintenance {Q4, 2025} 20](#_Toc174081384)

[☐ Deployment 20](#_Toc174081385)

[☐ Support 20](#_Toc174081386)

[☐ Updates 20](#_Toc174081387)

[☐ Monitor 20](#_Toc174081388)

[References 21](#_Toc174081389)

[Appendix {X} {Appedix Title} 24](#_Toc174081390)

[Title]

{Add Introduction}

Background on CNL and CTI {August 2024}

**Objective:** Establish a foundational understanding of Controlled Natural Languages (CNL) and Cyber Threat Intelligence (CTI), bridging the knowledge gap between cyber specialists and linguistics.

## Introduction to CNLs

**Objective:** Explain the concept of Controlled Natural Languages to non-linguists.

Controlled Natural Languages (CNLs) are simplified versions of natural languages designed to enhance clarity and eliminate ambiguity. They adhere to strict rules and limited vocabulary, making them easier for both humans, particularly non-native speakers, and machines to understand. CNLs are particularly useful in fields like cybersecurity, where precise and unambiguous communication is crucial (Haralambous, 2024).

### Definition and Purpose

**Objective:** Describe what a CNL is and how it differs from natural language.

CNLs are a subset of natural languages that follow specific rules to ensure consistency and clarity. Unlike natural languages, which can be complex and open to multiple interpretations, CNLs limit vocabulary and grammar to eliminate ambiguity. This makes CNLs ideal for contexts where precise communication is essential, such as technical documentation and formal specifications.

### Benefits

**Objective:** Discuss the advantages of using CNLs, such as reducing ambiguity, improving clarity, and facilitating automated processing.

CNLs offer several benefits:

* **Reduced Ambiguity:** By adhering to strict rules, CNLs minimize the risk of misunderstandings and misinterpretations.
* **Improved Clarity:** Simplified vocabulary and grammar ensure that messages are clear and easily understood.
* **Facilitated Automated Processing:** The structured nature of CNLs makes it easier for software to parse and interpret the language, enabling automated analysis and processing.

### Applications

**Objective:** Highlight the use cases of CNLs in various domains such as technical documentation, legal texts, and compliance.

CNLs are used in a variety of domains:

* **Technical Documentation:** CNLs ensure that technical instructions and specifications are clear and precise, reducing the likelihood of errors (Rese et al., 2023).
* **Legal Texts:** In legal contexts, CNLs help eliminate ambiguity, ensuring that legal documents are interpreted consistently.
* **Compliance:** CNLs are used to create clear and unambiguous compliance documents, facilitating adherence to regulations and standards.

In cybersecurity, CNLs can be particularly valuable for documenting threats, vulnerabilities, and incident responses. By using CNLs, cyber specialists can ensure that critical information is communicated clearly and accurately, enhancing collaboration and improving the effectiveness of threat mitigation strategies.

## Introduction to CTI and STIX

**Objective:** Provide background information on CTI and STIX for linguistics.

Cyber Threat Intelligence (CTI) and Structured Threat Information eXpression (STIX) are foundational concepts in cybersecurity, essential for understanding how cyber threats are identified, analyzed, and mitigated. This section aims to elucidate these concepts for linguistics, highlighting the importance of precise language in this domain.

### Cyber Threat Intelligence (CTI)

**Objective:** Define CTI and its importance in cybersecurity operations.

Cyber Threat Intelligence (CTI) involves the collection, analysis, and dissemination of information about cyber threats. CTI helps organizations understand the threats they face, providing insights that inform defense strategies and response actions. By analyzing threat data, cybersecurity professionals can anticipate and mitigate attacks more effectively, ensuring the security of information systems.

### Indicators of Compromise (IOC)

**Objective:** Introduce atomic and behavioral IOCs

Indicators of Compromise (IOCs) are pieces of information that suggest a system may have been breached. They are crucial for detecting and responding to cyber threats. There are two main types of IOCs:

* **Atomic IOCs:** These are simple, unambiguous indicators such as IP addresses, file hashes, or domain names. Machines can readily extract, understand, and act on atomic IOCs due to their straightforward nature. Regular grammars define atomic IOCs.
* **Behavioral IOCs:** These are more complex indicators that describe patterns of behavior, such as sequences of actions or tactics used by attackers. While atomic IOCs are easy for machines to process, behavioral IOCs are more challenging because they require understanding context and patterns, which are less structured and harder to codify.

### Structured Threat Information eXpression (STIX)

**Objective:** Introduce the STIX framework, its components, and its role in standardizing CTI.

Structured Threat Information eXpression (STIX) is a standardized language for representing CTI. It enables organizations to share threat information in a consistent and structured manner (Jordan et al., 2021). STIX includes components such as:

* Atomic IOCs
  + **Observables:** Descriptions of specific properties like IP addresses or file hashes.
  + **Indicators:** Patterns that suggest malicious activities.
  + **Incidents:** Descriptions of actual security events.
* Behavioral IOCs, also known as TTPs (Tactics, Techniques, and Procedures): Detailed descriptions of methods used by attackers.

STIX facilitates interoperability between different security tools and organizations, allowing for more effective sharing and utilization of threat intelligence.

## Intersection of CNL and CTI

**Objective**: Explain the intersection of CNL and CTI, emphasizing the need for precise language in documenting and analyzing cyber threats.

The intersection of Controlled Natural Languages (CNL) and Cyber Threat Intelligence (CTI) is critical for improving the clarity and precision of threat documentation. While machines can readily extract, understand, and act on atomic IOCs due to their simplicity, they struggle with behavioral IOCs and more complex STIX descriptions. This is where CNLs come into play.

CNLs can standardize the language used in behavioral IOCs and STIX descriptions, making it easier for both humans and machines to understand and process threat information. By using CNLs, cybersecurity professionals can reduce ambiguity and improve the accuracy of threat descriptions. This, in turn, enhances the ability of automated systems to analyze and act on CTI, leading to more effective threat detection and response.

In conclusion, the integration of CNL into CTI practices can significantly improve the effectiveness of cybersecurity operations by ensuring that threat information is clear, precise, and actionable.

Objectives and Scope {July 2024}

**Objective:** Establish the purpose and application domains of STIX-D.

## Purpose:

The STIX-D CNL aims to provide a ‘common tongue’ in which to communicate precise and unambiguous information in STIX description fields so that humans and machines can more readily understand and act on this information.

### Standardization Between {Dialects}

Many threat intelligence providers, such as MITRE, Crowdstrike, and Mandiant, write and share STIX descriptions. Each of these creators {speaks their own dialect}. Even within the same organization, {dialects} may differ between departments or purposes. For example, STIX descriptions differ noticeably between MITRE ATT&CK and MITRE CAPEC.

## Users and Requirements

**Objective:** Identify the primary users and their requirements.

## Tasks

**Objective:** Determine the specific tasks and contexts where STIX-D will be used.

## Scope

**Objective:** Define the scope and boundaries of STIX-D to ensure it meets the needs of its users.

Existing Languages and Standards {September 2024}

**Objective:** Leverage existing languages and frameworks to inform the development process.

## Existing STIX-Ds

**Objective:** Review existing STIX-Ds and their design principles.

## Domain Standards

**Objective:** Examine CTI-specific standards and terminologies (e.g., MITRE ATT&CK, STIX, etc.).

## Best Practices

**Objective:** Identify best practices and lessons learned from previous STIX-D implementations.

Corpus {September 2024}

**Objective:** Build a comprehensive corpus of source documents to inform the development of STIX-D.

## Source Documents

**Objective:** Gather a diverse set of documents relevant to the domain, including:

* Existing STIX descriptions
* MITRE ATT&CK framework
* Common Attack Pattern Enumeration and Classification (CAPEC)
* Narrative Cyber Threat Intelligence (CTI) reports

## Annotation and Analysis

**Objective:** Annotate and analyze the collected documents to identify common patterns, terminologies, and structures.

## Corpus Management

**Objective:** Ensure the corpus is well-organized and accessible for ongoing reference and analysis during the development of the CNL.

Lexicon {August 2024}

**Objective:** Create a foundational vocabulary for STIX-D.

## Key Concepts and Terms

**Objective:** Identify key concepts and terms relevant to the CTI domain.

## Definitions

**Objective:** Define precise meanings for each term to avoid ambiguity.

## Consistency

**Objective:** Ensure consistency in terminology across STIX-D.

Grammar Rules {Q4, 2024}

**Objective:** Define the syntactic and semantic rules for constructing sentences in STIX-D.

## Syntax

**Objective:** Develop a set of grammar rules that govern how words and phrases can be combined.

## Semantics

**Objective:** Ensure the rules promote clarity and prevent ambiguity.

## Testing

**Objective:** Test the grammar rules with example sentences to ensure they produce the desired results.

Semantic Framework {Q1, 2025}

**Objective:** Develop a framework for understanding and processing the meanings of sentences.

## {ADD SECTION HEADING}

**Objective:** Define the relationships between different terms and concepts.

## {ADD SECTION HEADING}

**Objective:** Implement mechanisms for resolving ambiguities and ensuring semantic consistency.

## Testing

**Objective:** Develop methods for validating and verifying the semantic content of STIX-D sentences.

Tools and Resources {Q2, 2025}

**Objective:** Provide users with tools to effectively use and interact with STIX-D.

## Authoring Tools

**Objective:** Develop authoring tools to help users create STIX-D-compliant texts.

## Parsing Tools

**Objective:** Implement parsers to {WHAT} of STIX-D sentences.

## Validators

**Objective:** Implement validators to check the correctness of STIX-D sentences.

Testing and Evaluation {Q2, 2025}

**Objective:** Ensure STIX-D meets its design objectives and user requirements.

## {Add Heading}

**Objective:** Test STIX-D with real-world use cases and scenarios.

## {Add Heading}

**Objective:** Gather feedback from users and subject matter experts.

## Improve

**Objective:** Iterate on the design based on feedback and testing results.

Documentation and Training {Q3, 2025}

**Objective:** Ensure users understand how to use STIX-D correctly.

## Documentation

**Objective:** Develop comprehensive documentation, including a user manual and reference guide.

## Training Materials

**Objective:** Offer training sessions and resources to educate users on STIX-D.

## {Add Heading}

**Objective:** Create examples and case studies to illustrate proper usage of STIX-D.

Deployment and Maintenance {Q4, 2025}

**Objective:** Ensure STIX-D remains effective and relevant over time.

## Deployment

**Objective:** Deploy STIX-D and associated tools to the user community.

## Support

**Objective:** Provide support and updates based on user feedback and evolving requirements.

## Updates

**Objective:** Provide updates based on user feedback and evolving requirements.

## Monitor

**Objective:** Monitor the usage and effectiveness of STIX-D, making adjustments as needed.

References

{ *List of articles and any books cited* – an alphabetized list of the sources that are cited in the paper (by last name of the first author of each source). Each reference should follow specific APA guidelines regarding author names, dates, article titles, journal titles, journal volume numbers, page numbers, book publishers, publisher locations, websites, and so on (for more information, please see the [Citing References in APA Style](https://psychology.ucsd.edu/undergraduate-program/undergraduate-resources/academic-writing-resources/writing-research-papers/citing-references.html) page of this website). }

{*Graphs and data* (optional in some cases) – depending on the type of research being performed, there may be Tables and/or Figures (however, in some cases, there may be neither). In APA style, each Table and each Figure is placed on a separate page and all Tables and Figures are included after the References. Tables are included first, followed by Figures. However, for some undergraduate papers, Tables and Figures may be embedded in the text (depending on the instructor’s policies; for more details.}

Figure

Sample Figure

A picture containing screenshot, black, diagram, graphics

Description automatically generated

Notes. {Enter any notes here or delete this line.}

Appendix {X}  
{Appedix Title}

{Sometimes authors wish to include material that supplements the paper’s content but that would be distracting or inappropriate in the text of the paper. Such material can often be included in an appendix, which is included in the print and electronic versions of the article, or in supplemental materials (see Section 2.15), which are available in an online-only supplemental archive that the publisher maintains.

Include an appendix only if it helps readers understand, evaluate, or replicate the study or theoretical argument being made. Be sure that all relevant ethical standards have been followed for materials placed in the appendices, including copyright attribution, accurate representation of data, and protection of human participants (e.g., as the standards apply to images or videos of identifiable people; see Sections 1.18 and 12.17).

In general, an appendix is appropriate for materials that are relatively brief and easily presented in print format. Some examples of material suitable for an appendix are (a) lists of stimulus materials (e.g., those used in psycholinguistic research); (b) instructions to participants; (c) tests, scales, or inventories developed for the study being reported; (d) detailed descriptions of complex equipment; (e) detailed demographic descriptions of subpopulations in the study; and (f) other detailed or complex reporting items described in Chapter 3. Student papers may include appendices.}