Finding Genealogy Facts
with
Linguistic Analysis

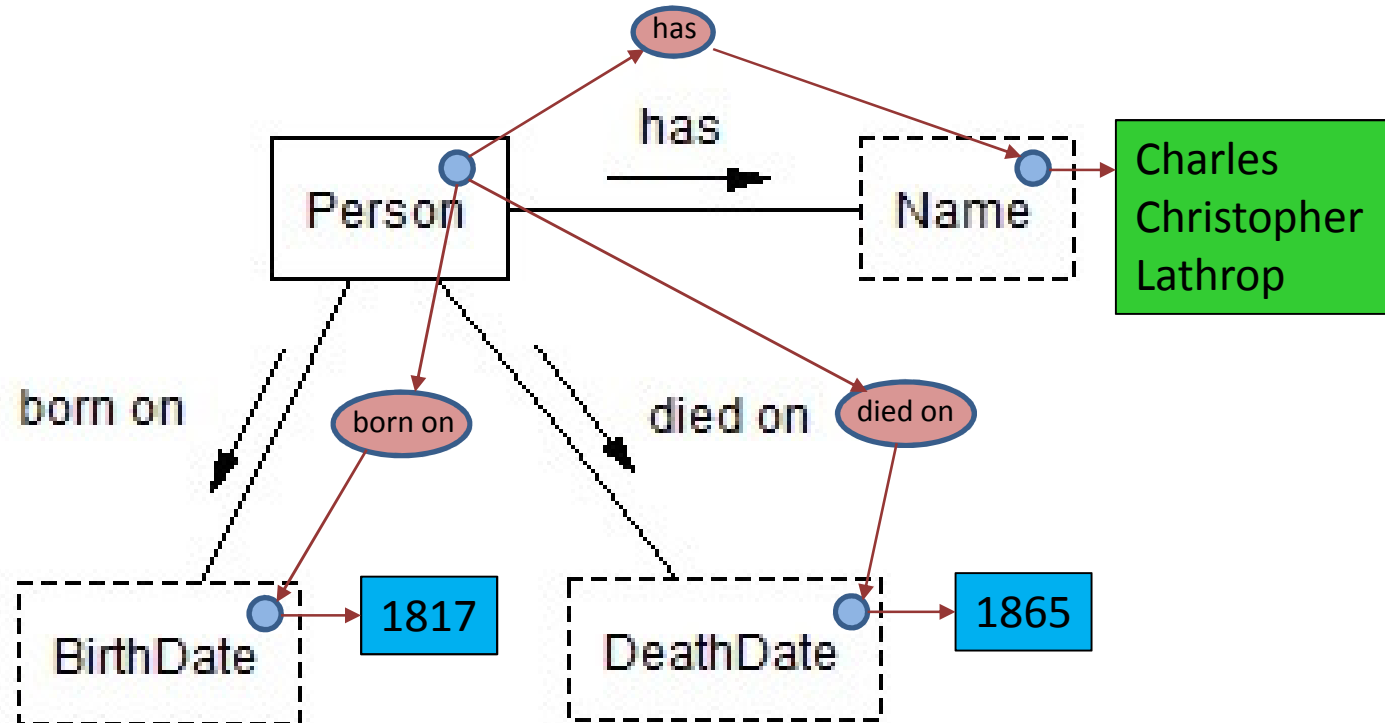Peter Lindes, Deryle W. Lonsdale, David W. Embley

Brigham Young University

Semi-Structured Text

243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865, son of Mary Ely and Gerard Lathrop; m. 1856 Mary Augusta Andruss, 992 Broad St., Newark, N. J., who was b. 1825, dau. of Judge Caleb Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4, 1898. The funeral services were held at her residence on Monday, Nov. 7, 1898, at half-past two o'clock P. M. Their children:

1. Charles Halstead, b. 1857 d. 1861.
2. William Gerard, b. 1858, d. 1861.
3. Theodore Andruss, b. 1860.
4. Emma Goble, b. 1862.

Miss Emma Goble Lathrop, official historian of the New York Chapter of the Daughters of the American Revolution, is one of the youngest members to hold office, but one whose intelligence and capability qualify her for such distinction.
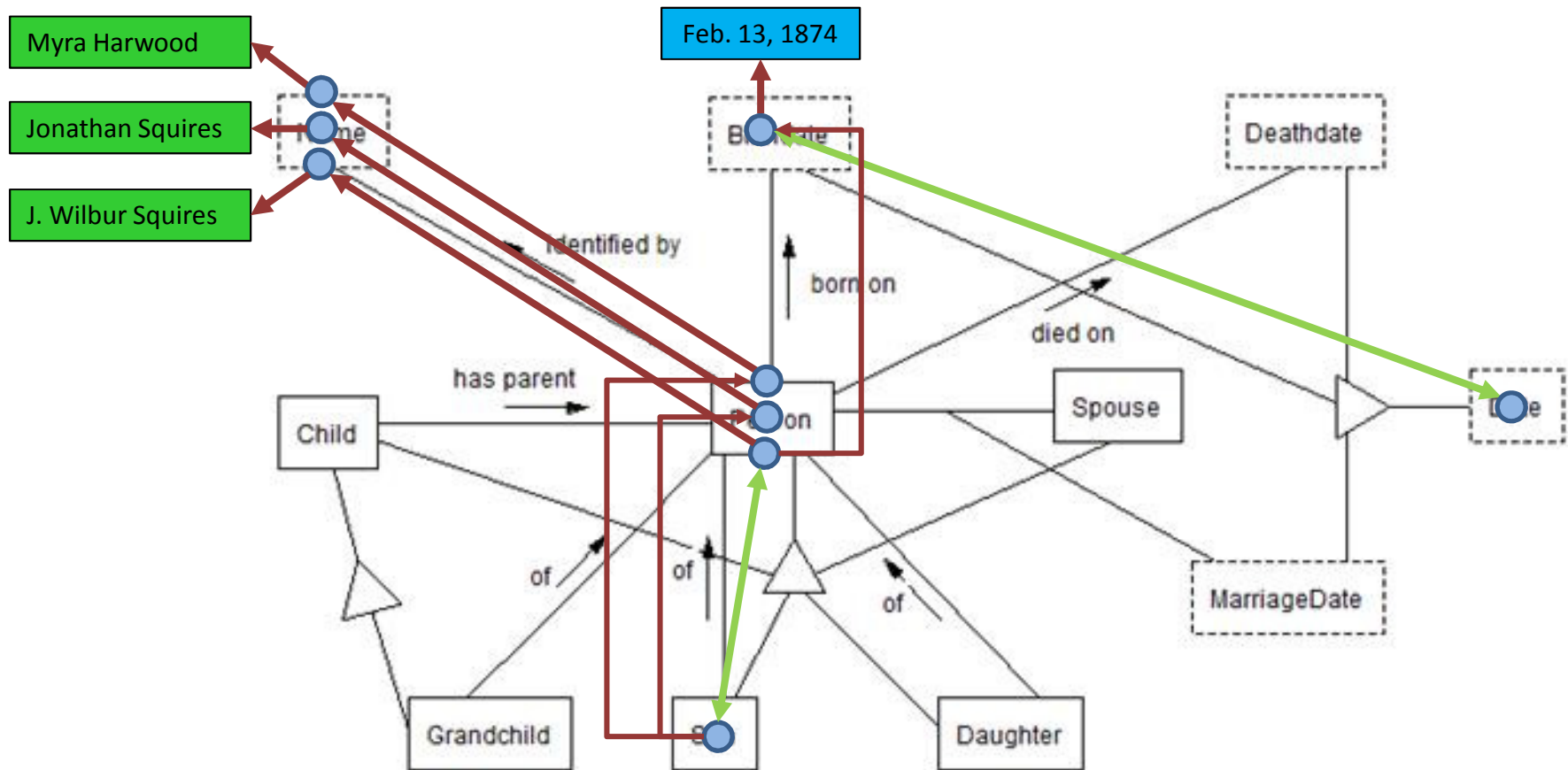
A Simple Ontology

Unstructured Text
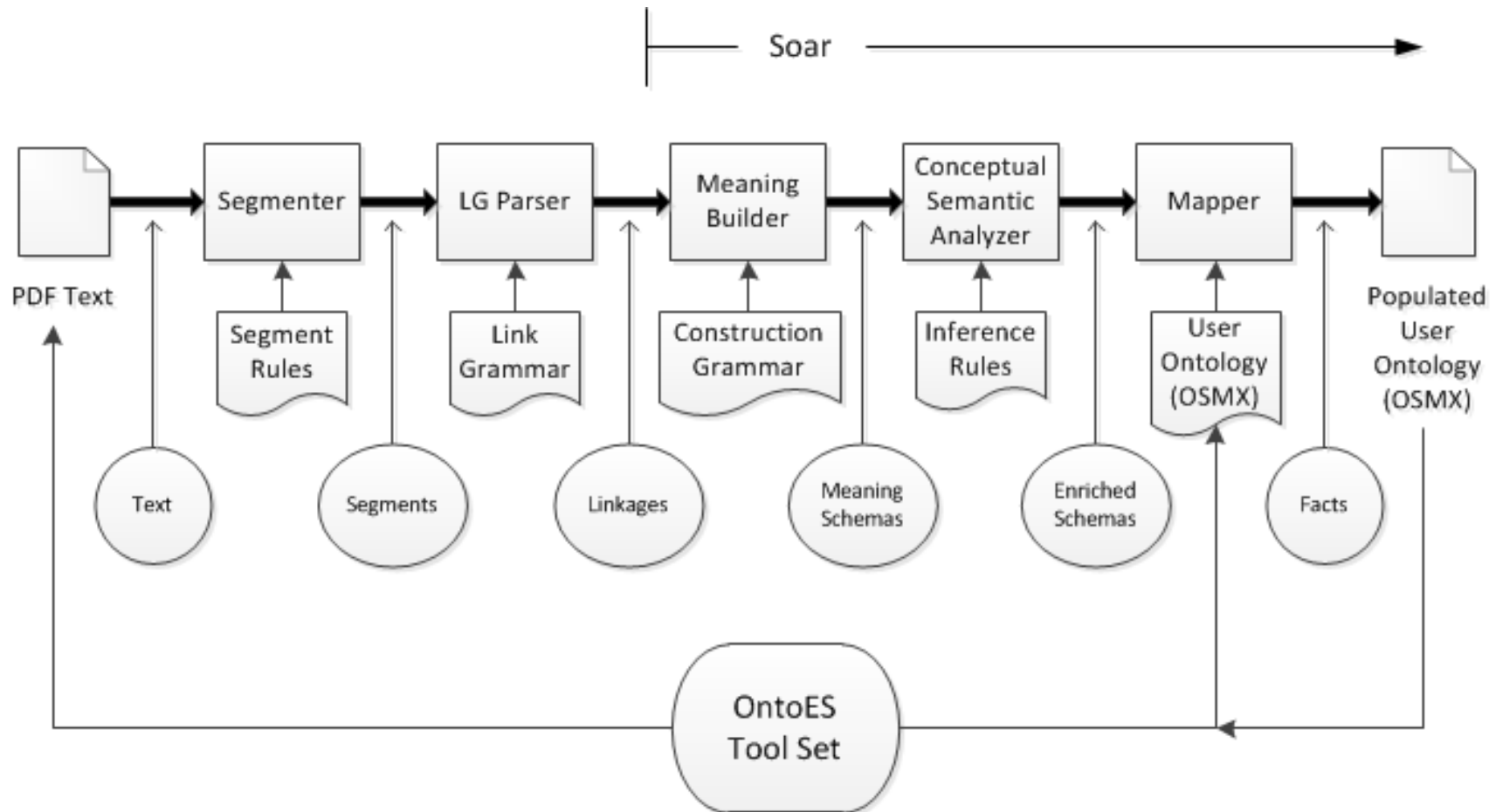

Children of JAMES HARWOOD, No. 103.

229. M——, born July 26, 1835, in Eden, Vt. She married Elijah Spencer, Dec. 25, 1851. They had five children: Arvilla, born in 1852, is not living; Mariette, born Dec. 25, 1854, married Jonathan Snyder, have a family; Leverett, born Feb. 6, 1857, married Cora Smith, Nov. 2, 1879, had two children, Perry F. and Ida I. Leverett died May 21, 1910; Rosa E., born Jan. 13, 1860, married Emmett Byers, and have children; and Harrison, born about 1862, is not living. Elijah Spencer died in the Union army in 1863, and his widow married Jonathan Squires, who was born in Ohio, July 25, 1823, by whom she had one son, J. Wilbur, born June 16, 1865, in DeKalb county, Ind., married Cora M. Thomas, Aug. 24, 1887, they reside in St. Joseph, Mich., five children. Mrs. Myra Squires died in Allen county, Ind., Feb. 13, 1874.

A More Complex Ontology

OntoSoar Architecture

Two Cases

- ## Case 1:

  Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865, son of Mary Ely and Gerard Lathrop ;

- ## Case 2:

  his widow married JONATHAN SQUIRES, who was born in Ohio, July 25, 1823, by whom she had one son, J. Wilbur, born June 16, 1865, in DeKalb county, Ind.

Linguistic Analysis – Case 1A

Charles Christopher Lathrop, …, b. 1817, d. 1865,

Linguistic Analysis– Case 1B

Charles Christopher Lathrop, … , son of Mary Ely and Gerard Lathrop ;

Extracting Meaning – Case 1

Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865, son of
Mary Ely and Gerard Lathrop ;

Linguistic Analysis– Case 2

his widow married JONATHAN SQUIRES, …, by whom she had
one son, J. Wilbur, born June 16, 1865, … .

Extracting Meaning – Case 2

his widow married JONATHAN SQUIRES, who was born in Ohio, July 25, 1823, by whom she had one son, J. Wilbur, born June 16, 1865, in DeKalb county, Ind.
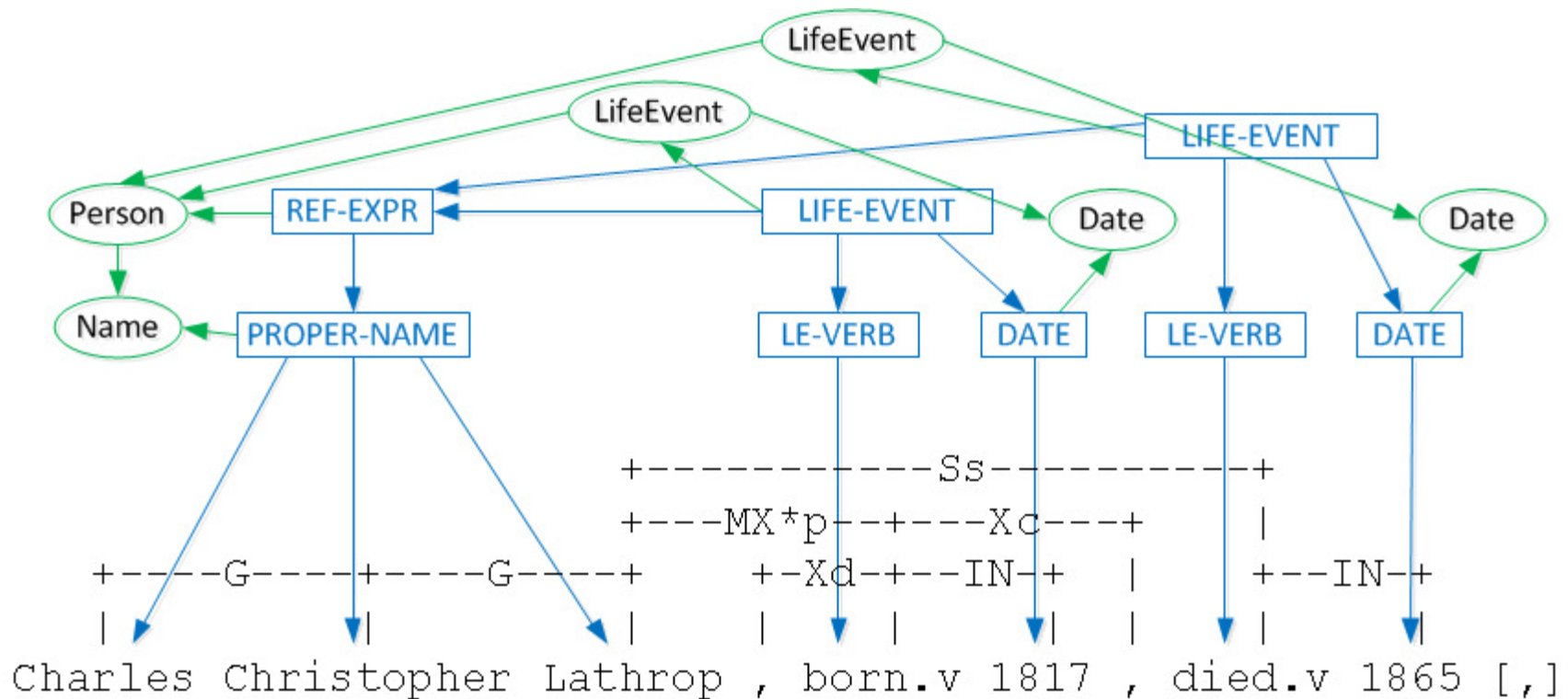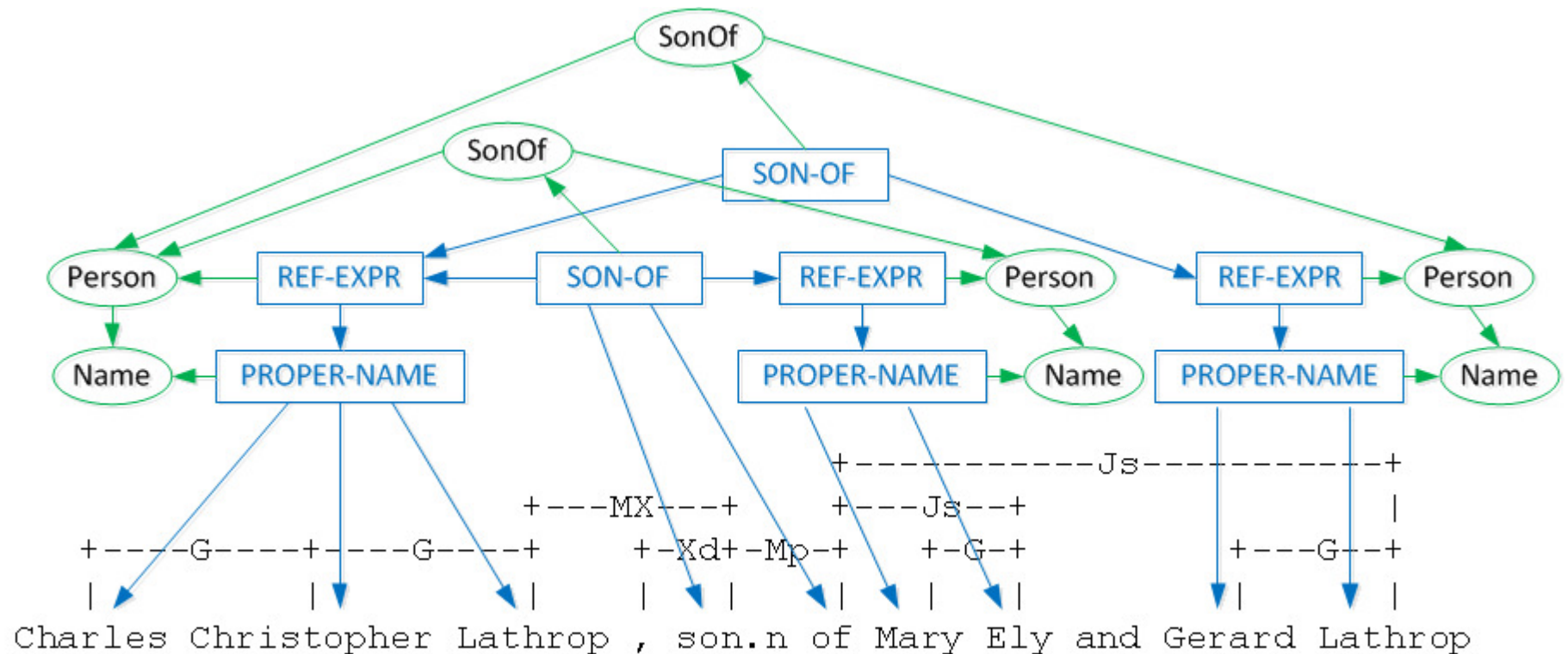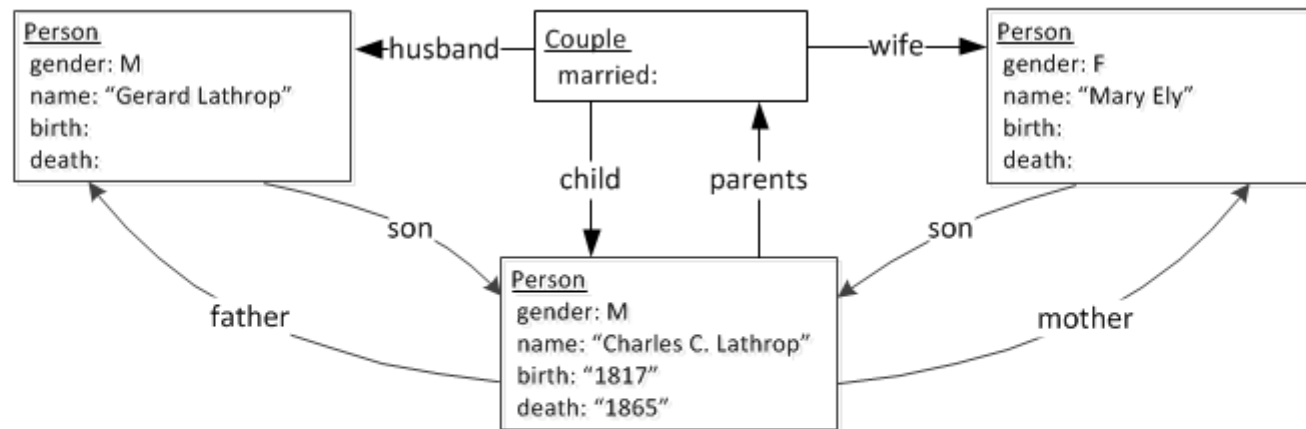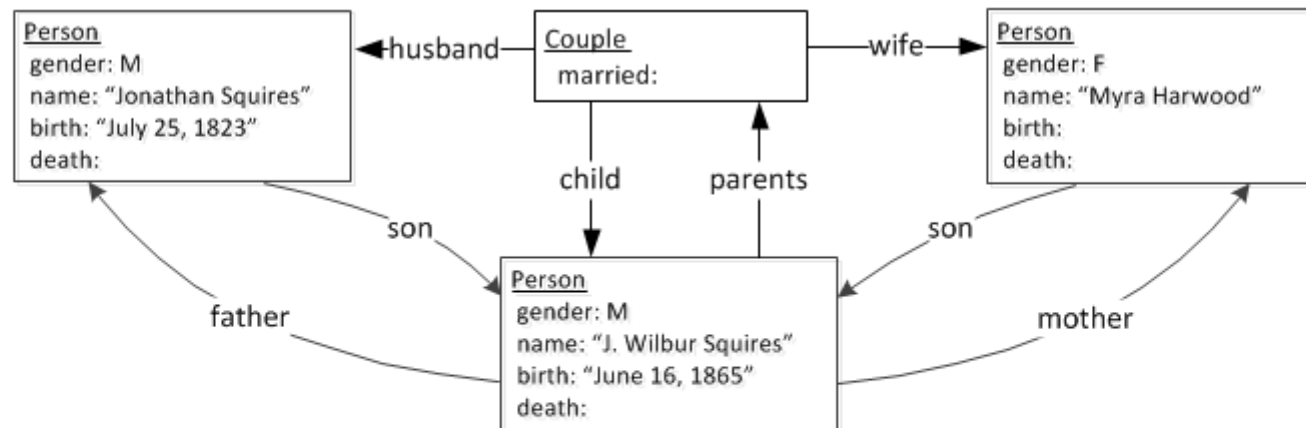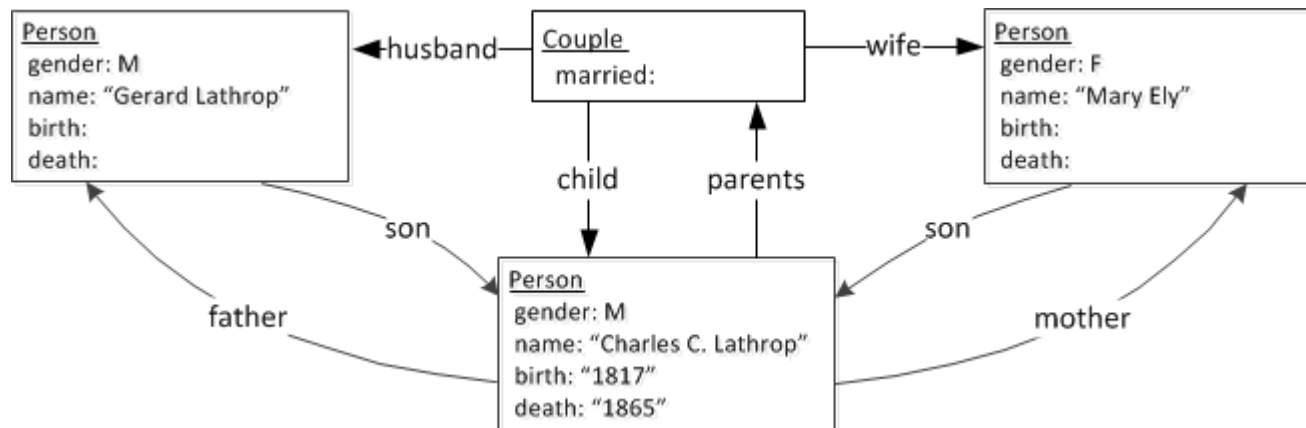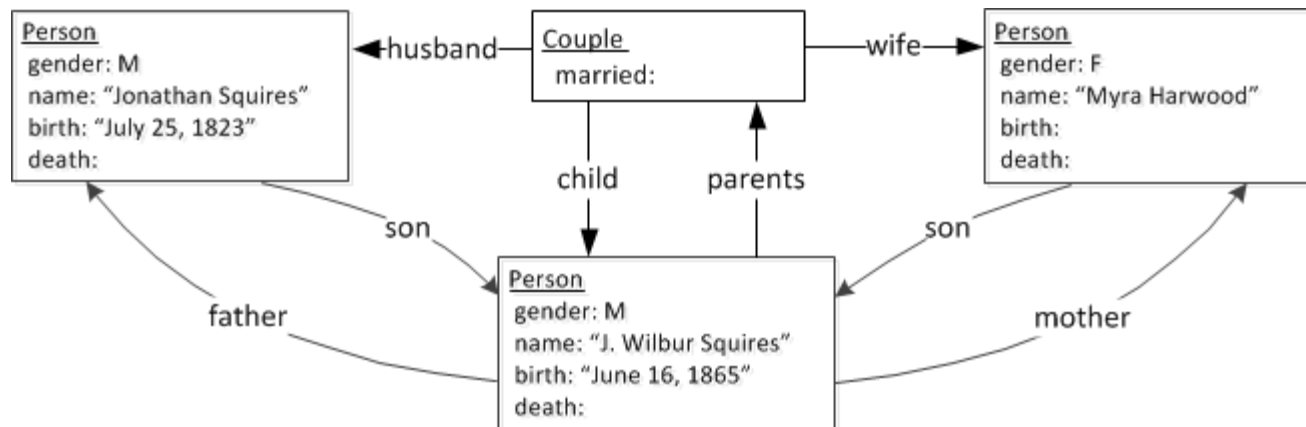
Meaning Structures Compared

Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865, son of Mary Ely and Gerard Lathrop ;

his widow married JONATHAN SQUIRES, who was born in Ohio, July 25, 1823, by whom she had one son, J. Wilbur, born June 16, 1865, … .

Evaluation

Conclusions

Accomplishments

- Preprocessor/Segmenter
- Adapted LG Parser for genealogy text
- Built a construction grammar meaning analyzer
- Mapped meaning schemas to user-defined ontologies
- Integrated all into a single processing pipeline
- Partially implemented – works well so far

Future work

- Complete implementation
- Evaluate on a broader corpus
- Explore using a named entity recognizer
- Build in much more linguistic knowledge
- Research how to make the system learn linguistic knowledge