What questions can humanists and social scientists find answers to with the computational tools in the NLP Suite?

Contents

Humanists' and social scientists' questions	Computer scientists' answers and available NLP Suite Tools
What's in your documents? A broad view	
What does your corpus talk about? What are the **topics**?	*Under CORPUS analysis tools select:* • Topic modeling (via MALLET & Gensim)
Which words come together in texts? Are women/girls mostly associated with adjectives or nouns of physical beauty, family relations, and low-paying occupations in the care industry (waiters, teachers, nurses)? Is the opposite true for men and boys?	*Under CORPUS/DOCUMENT analysis tools select:* • Word embeddings (Word2Vec) (via BERT & Gensim) • CoNLL table analyzer
Is there a way to get basic **statistics about your *corpus***, i.e., the set of documents you are studying? For example, the number of documents, the number of words per document, the sentence length per document?	*Under CORPUS/DOCUMENT analysis tools select:* • Corpus/document(s) statistics (Sentences, words, lines)
What's in your documents? A closer look	
Syntactic relations between words	
Can computational tools tell you **which words in a text are nouns, verbs, or adjectives**? Yes. There are basic tools called **parsers** that do precisely that, and more and with a high degree of accuracy (over 90%). They	The NLP Suite has several different options for parsers: spaCy, Stanford CoreNLP, Stanza, each with slightly
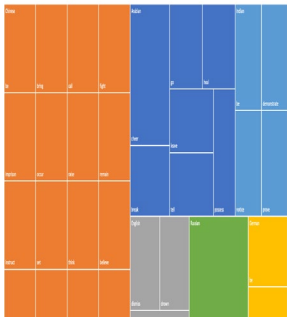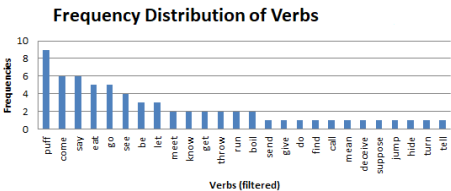
can even give the syntactical relation between words, whether verbs are in the infinitive form, gerundive, past, present, passive or active, whether nouns are singular or plural subjects or objects, and, again, active or passive.	different characteristics of accuracy or speed. *Under CORPUS/DOCUMENT analysis tools select:* • Parsers & annotators (BERT, CoreNLP, spaCy, Stanza)
Can they even tell you **which adjectives are used for which nouns**? Like in a study of folktales, which adjectives are used for princesses and princes? Definitely. Parsers produce in output a CoNLL table that contains all the information required to answer those questions.	In the NLP Suite, the CoNLL table analyzer algorithms will allow you to explore these questions. *Under CORPUS/DOCUMENT analysis tools select:* • CoNLL table analyzer - Search the CoNLL table
There is a problem in linguistics known as **pronominal anaphora** where in such a sentence as "John hurried home. He said he was ill," where both "he" pronouns (the anaphoric expression) refer to "John" (the antecedent). Can computational tools link pronouns to their nouns (anaphoric expressions to their antecedents)? In NLP, this problem is known as **coreference resolution**. Unfortunately, the accuracy of these algorithms is not as high as for parsers, even for neural network approaches (around 70%?).	There are two points of entry to coreference resolution the NLP Suite and with several different options available. *Under CORPUS/DOCUMENT analysis tools select:* • CoreNLP annotator - coreference (pronominal) • Parsers & annotators (BERT, CoreNLP, spaCy, Stanza)
Subject-Verb-Object (SVO) extraction and visualizations (the 5Ws of narrative)	
Who does what, where, when, and why? Can NLP help to identify the **5Ws of stories**?	*Under CORPUS/DOCUMENT analysis tools select:*

	• SVO (Subject-Verb-Object) extractor & visualization

<table>
<tr><td colspan="2" align="center">Semantic relations between words</td></tr>
<tr>
<td>Which words come together in texts? Are women/girls mostly associated with adjectives or nouns of physical beauty, family relations, and low-paying occupations in the care industry (waiters, teachers, nurses)? Is the opposite true for men and boys?</td>
<td>Under CORPUS/DOCUMENT analysis tools select:

• Word embeddings (Word2Vec) (via BERT & Gensim)</td>
</tr>
<tr>
<td>Do nouns and verbs cluster together in specific groups (e.g., verbs of movement or verbs of communication)?</td>
<td>Under CORPUS/DOCUMENT analysis tools select:

• WordNet</td>
</tr>
<tr><td colspan="2" align="center">People, organizations, time, space: Zooming into the texts</td></tr>
<tr>
<td>Are there male or female characters in the writings? Where do they appear in the texts? At the beginning, middle, end, dispersed throughout? Do they come together? Is it a man's world or a woman's world?</td>
<td>There are several tools in the NLP Suite that allow you to find female/male characters in a text, from the Stanford CoreNLP gender annotator to the various databases with first names tagged as female or male (e.g., US Social Security name database and others).

Under CORPUS/DOCUMENT analysis tools select:

• CoreNLP annotator - gender (male & female names; via CoreNLP and dictionaries)</td>
</tr>
<tr>
<td>Are there people, organizations, temporal expressions, or locations mentioned in the texts?</td>
<td>Under CORPUS/DOCUMENT analysis tools select:</td>
</tr>
</table>

	• NER (Named Entity Recognition) annotator
If documents in your corpus mention proper names (people, organizations, locations) can you find **information on the web** automatically? For instance, when and where they were born, their occupation.	*Under CORPUS/DOCUMENT analysis tools select:* • Knowledge graphs: DBpedia & YAGO
The world of sentiments and emotions	
Which **sentiments and emotions** do writers express in their writing? Positive, negative? How do they express these emotions rhetorically and linguistically?	The NLP Suite provides several options to address these questions. To find all the options *Under CORPUS analysis tools select:* • Sentiments/emotions (ALL options GUI) The GUI will give you access to all the various options available in the NLP Suite to find answers to those questions. You can also select • Style analysis (ALL options GUI) Then tick the Vocabulary analysis option and select: Punctuation as figures of pathos (?!) Or the Repetition tools (repetition is also a figure of pathos). Sentiment analysis (neural network/tensor options:

	BERT, spaCy, Stanford CoreNLP, Stanza) Sentiment analysis (dictionary options: ANEW, hedonometer, SentiWordNet, VADER) You can also use YAGO's emotion ontology class or WordNet (NOUN feeling VERB emotion) to search for emotion words used in the texts
And when a single story is viewed in the context of hundreds or thousands of other stories, do these stories cluster together in specific **story shapes**, as Kurt Vonnegut would have it? A man in a hole, from rags to riches or riches to rags? Happy ending or a bummer? Comedy or tragedy?	*Under CORPUS analysis tools select:* • Shape of stories
Writing style	
Which **style** do different authors use in their writing? Can the NLP Suite help highlight general characteristics of an author's style?	The NLP Suite provides a wide range of options to find answers to these questions. *Under CORPUS/DOCUMENT analysis tools select:* • Style analysis (ALL options GUI)
Are there words and combinations of words that appear more frequently in texts? How do they different across different authors? N-grams, Co-occurrences, *N-grams/co-occurrences viewer* Computationally, these objects are known as n-grams: unigram for single words, bigrams for combinations of two words, trigrams for three words, etc. (although it generally makes little sense to go beyond trigrams).	*Under CORPUS/DOCUMENT analysis tools select:* • N-grams (word & character) • N-grams/co-occurrences viewer

Is there a consistent use of passive and active verb forms, of present and past verb tenses, of some subjects always portrayed in active or passive forms, acknowledging or denying agency? *Verb modality, tense, voice* *Nominalization*	*Under CORPUS/DOCUMENT analysis tools select:* • CoNLL table analyzer - Search the CoNLL table • Nominalization
How **readable is a text**? How **complex are the sentences**? *Sentence complexity* *Text readability*	*Under CORPUS/DOCUMENT analysis tools select:* • Sentence/text readability (via textstat) • Sentence complexity
How is the text vocabulary? Are there **unusual words**, **misspelled words**? Are there clear patterns in the **first and last few sentences** of documents? More nouns, more verbs… Are there **repetitions**? How rich is an author's vocabulary? *Vocabulary* Unusual words/misspelled words Short/long words, vowel words Vocabulary richness Abstract/concrete vocabulary Objective/subjective writing In which language are texts written in?	*Under CORPUS/DOCUMENT analysis tools select:* • Style analysis (ALL options GUI) Then tick the Vocabulary analysis option and explore the many options available.
Who wrote the text? Man or woman?	*Under CORPUS/DOCUMENT analysis tools select:* • Who wrote the text? Man or woman? (via Gender guesser)
Is there *dialogue in the text*? And who does the talking? *Is there dialogue in the text?*	Stanford CoreNLP, uniquely, has a special quote annotator that provide answers to these questions. *Under CORPUS/DOCUMENT analysis tools select:*

	- Parsers & annotators (BERT, CoreNLP, spaCy, Stanza)
Visualizing words	
How can you get those pretty pictures of words in different colors and sizes? *Wordclouds* 	This type of image is called a wordcloud, easily done in the NLP Suite and with a wide range of options. *Under CORPUS/DOCUMENT analysis tools select:* - Wordclouds (ALL options GUI)
How can you visualize relations between people or words? *Sankey plot* 	This type of plot is called a Sankey plot. *Under Visualization tools select:* - Sankey flowchart (Plotly) (Open GUI)
How can you visualize relations between people or words (e.g., Subject-Verb-Object)? *Network graphs* 	This type of plot is called a network graph and you can display one easily in the NLP Suite (and some NLP tools, for instance SVO, export a network graph automatically). *Under Visualization tools select:* - Network graphs (Gephi) (Open GUI)
Is it possible to visualize gendered names (male or female) in different colors to have an immediate view of where male/female names occur in the text?	*Under Visualization tools select:*

HTML files ‹⊕›A Spool of Blue Thread _Anne Tyler_Rebecca Pepper Sinkler_02-13-2015.txt‹⊕› Picture a middle-class American family, four generations, living in the suburbs. The "founder" is an upstart with a murky past, a hard-working construction guy in a bleak marriage. Next in line are his son and the son's wife — he's a chip off the old patriarch, inherits the construction business, soldiers on. She's the maddening, meddling family fixer-upper, determined that her four children will, against all odds, tolerate if not love one another. And then there are the children, at midlife and with children of their own: a couple of daughters who annoy each other, an "acquired" son (the product of a "crazy arrangement") and of course the obligatory black sheep, Denny, the prodigal son who surfaces intermittently to provide the drama, which, after all, is his job — in the family as in the plot of Anne Tyler's graceful and capacious new novel, "A Spool of Blue Thread." So what else is new? Give or take a few details, this appended blended fouled-up family could be any of ours. That makes it cliché territory, risky for an ambitious novelist. It's also quintessential Anne Tyler, as well as quintessential American comedy. Tyler has a knack for turning sitcom situations into something far deeper and more moving. Her great gift is playing against the American dream, the dark side of which is the falsehood at its heart: that given hard work and good intentions, any family can attain the Norman Rockwell ideal of happiness — ordinary, homegrown happiness. Happiness indeed eludes the Whitshank family of Baltimore. That's where Junior, the founder, begins by constructing a house, a place that becomes the love of his life, as well as that of his son, Red, and eventually some of the other children and grandchildren. The house is initially commissioned by a rich businessman, but Junior, who is hired to build it, sets his heart on the place the minute he sees the blueprints, crafts it to his own high standards and eventually acquires it by some mild chicanery. The house is no Satpen's Hundred, no Manderley; "more, let's say, a family house. A house you might see pictured on a thousand-piece jigsaw puzzle, plain-faced and comfortable" with a "wonderful full-length porch." It projects ordinariness, togetherness, timelessness. And it has the power to draw the family back to it over the generations. Togetherness and ordinariness are mixed blessings. Junior's family is anything but ordinary. Early on, he's entrapped by a 13-year-old seductress (half his age) who bides her time and eventually gets her man. (The Whitshanks are champion time-biders.) Junior and his wife in turn give birth to Redcliffe (boy) and Merrick (girl), their highfalutin names bestowed by Junior, who, much as he may resent his snooty neighbors, apes their ways. (Tyler has fun with Merrick, the kind of girl who steals her best friend's beau and unstays off into the upper echelons of Baltimore society, where people have names like Pookie and Pixie.) While Merrick escapes, her brother, Red, takes on the family construction business and a wife named Abby. Alas, poor Abby! She's a staple of Tyler fiction: the self-denigrating self-aggrandizing embarrassment of a mother you hope will never show up at your school function. It's Abby, Red and their children we get to know best as the family slowly disintegrates. They're first met as the novel opens and Denny calls home to tell Red and Abby he's gay. (He's not.) "What the hell," Red snorts, whereupon Denny hangs up and Abby wigs out. The ensuing scene hilariously captures small fractures, now grown into ruts, in an old, well-worn marriage. Denny is a walking catastrophe: college dropout, erstwhile stoner, irregularly employed, an unmarried and truant father, sometimes homeless, usually out of reach. No one in the family, not his parents or his two sisters and younger brother, can cope with his existence, let alone his occasional homecomings. His older sister, Amanda, puts it best when, in a moment of exasperation, she confronts him with a litany of domestic crimes and misdemeanors. It's an inventory every family with a black sheep knows all too well — neglected birthdays, absence at anniversaries and weddings and family reunions, general insufferable narcissism — but the bill of wrongs finally names the real source of his siblings' pain: "But most of all, Denny, most of all. I will never forgive you for consuming every last little drop of our parents' attention and leaving nothing for the rest of us." Denny is one of Tyler's most endearing miscreants. He's surely unforgivable, but she has a soft spot for him and so she wins us over. When Red starts going deaf and then has a heart attack, when Abby's brain starts to "jump the track," the family converges to take over their care. Predictably, the "good" son, dutiful Stem, not even a Whitshank by birth (his birth mother's identity a closely held secret), and Stem's judicious wife, Nora, move into the old house to manage the aging parents. But — atypically — bad boy Denny suddenly arrives to shoulder a share of the burden. Here Tyler concocts a noxious brew of ancient jealousies, resentments and guilt as the siblings vie for control. After a family slugfest that ends in some minor bloodshed and a kitchen full of "fluttering women and shocked, wide-eyed children," the truth surfaces. "I sure don't want to be in charge," Denny cries. "It's that Dad isn't satisfied with the rest of us. . . . All my life Dad has made me feel I didn't quite measure up. . . . Like I'm lame; I'm lacking." Was that how the prodigal son felt? Is that why he left home in the first place? And what about his stay-at-home brother? Seeing his father kill the fatted calf to welcome him, the angry, dutiful brother had every right to complain. You wonder where the mother was in Jesus' parable. If she was anything like Abby, she, like God, was squarely on the side of the lost sheep. For Tyler, the quality of mercy is anything but strained. By her own admission, she's a chronicler of sad failures and unhappy marriages. In novel after novel she predisposes her characters to crave the unattainable — parental love (in both directions), a sense of belonging (among your own and in the world), forgiveness, amnesty from familial wrongdoing, the comfort of home. And yet she's a comic novelist, and a wise one. The calamities she depicts are minor, after all, and her characters aren't the twisted, fearsome ones of much American fiction. Her take on her crazies and misfits seems to line up with that of Abby, who, recalling the demise of the Wicked Witch of the West in "The Wizard of Oz," concludes, "I really believe that most people who seem scary are just sad." Interestingly, what Abby recalls most strongly among the witch's last words isn't just her cry, "I'm melting, I'm melting," as we all remember it. Rather, it's the witch's lament: "Who would have thought a good little girl like you could destroy my beautiful wickedness?" In "A Spool of Blue Thread," as in her work in general, goodness trumps wickedness, there's no place like home and, yes, we can all go there again, even if it isn't that perfect place we long for.	• HTML annotator - dictionary, gender, DBpedia, YAGO, WordNet - (All options GUI)
Can you visualize locations mentioned in texts as geographic maps? *Pin maps heat maps*	*Under CORPUS/DOCUMENT analysis tools select:* • Geographic maps: From texts to maps You can also map locations from a list of locations in a csv file: • Geographic maps: Google Earth Pro You will find the same GIS tools under *Visualization tools.*
I have seen presentations with fancy charts like this that move around when you click different parts of the chart. Can it be done easily in the NLP Suite? *Sunburst pie charts*	This type of chart is called sunburst pie chart. It is a powerful interactive visualization tool that you can easily use in the NLP Suite. *Under Visualization tools select:* • Sunburst pie chart (Plotly) (Open GUI)
How can you produce a map of this kind (from a comparative analysis of folktales across 6 different countries)? *Treemap*	This type of map is called a treemap and you can easily create one in the NLP Suite. *Under Visualization tools select:*

	• Treemap (Plotly) (Open GUI)
	Under Visualization tools. • Animated time-dependent bar plot (Plotly) (Open GUI)
Can this type of chart be easily produced to visualize the frequency distribution of certain aspects of texts (e.g., verbs like in this case)? *Bar charts, pie charts, line charts* 	Every tool in the NLP Suite routinely produces in output a range of plots in either Excel or Plotly, as bar charts or pie charts. You can also access the visualization tools *Under Visualization tools select:* • Excel & Plotly charts