

# **New Approach Needed for Tools**

**Since an NLP-based tool cannot achieve better than 85–91% recall,**

**perhaps it is time to try other approaches to design a tool for a HT.**

**An examination of the RE and SE tools literature shows a number of promising approaches worth pursuing.**

# Naive Use Even Worse

**As Ryan [1993] observed, naive use of such a tool may**

- 1. worsen the analyst's workload — the analyst looks at the tool's output and then has to do the whole manual analysis anyway**  
  
**or**
- 2. lull the analyst with unjustified confidence in the tool's output.**

# Rethinking Any NLP-Based RE Tool

**If the tool cannot save the analyst work ...**

**by doing 100% of analysis, and ...**

**the analyst must manually analyze the whole document anyway, ...**

**it might be best to forgo the tool and ...**

**focus on doing the manual analysis very well.**

# Rethinking, Cont'd

**Preparing to do well might include getting a good night's sleep the night before!**

# How to Use an Imperfect Tool

**The second risk (lulling) of naive use of a tool with recall  $< 100\%$  suggests that the best time to use such a tool is *after* a best-effort manual analysis that is felt to have been as thorough as possible.**

# After Manual Analysis is Done

**Now, anything that the tool finds**

- 1. that the analyst overlooked or**
- 2. that prompts the analyst to find something he or she overlooked**

**is a low-cost bonus.**

# But ...

**But, if the user *knows* that a tool *will* be used later, then he or she may nevertheless fall into the trap of being lulled!**

# Another Source of Same Recommendation

**This recommendation is consistent with Dekhtyar *et al.*'s observation that ...**

**when asked to vet traces proposed by an automatic tracer, a category (c) tool, humans tended to decrease both the recall and precision of the traces.**

**Knowing that a tool was used made them sloppier.**



# Novices' Use of a Tool

**Kiyavitskaya *et al.* have shown in an experiment that a high-precision, low-recall tool for annotating laws helps novices achieve 96% recall relative to experts.**

**I guess that the high precision helped the novices learn what is right, so that each could use his or her intelligence correctly.**

# Experts' Use of Same Tool

**Experts did not participate in Kiyavitskaya *et al.*'s experiment.**

**My bet is that ...**

**Experts using the tool will find their recall deteriorating.**

**We need to test.**

# Another Idea

**When no tool can do analysis  $A$  with 100% recall, ...**

**but there is an algorithmically *identifiable* part of  $A$  that can be done with 100% recall by some tool  $T$ , then ...**

**it might be useful to build  $T$  and let it do what it can, ...**

**so that the analyst can focus on only the part of  $A$  that cannot be done with 100% recall.**

# The Key of the Idea

**The key here is that the tool's and the human's parts of *A* are algorithmically identifiable, and ...**

**the tool's and the human's parts of *A* together are *all* of *A*.**

**So that the analyst can *really* ignore the tool's part of *A*, and thus can *really* focus on the human's part of *A*.**

# **SREE, An Example of Idea**

**Tjong's SREE, a category (a) ambiguity finding tool, finds ...**

**only those potential ambiguities that are identifiable by a lexical scanner.**

**It leaves all other ambiguities to be found manually.**

# Use of SREE

**SREE finds *all* potential instances of the “only” ambiguity by finding each sentence with the word “only”.**

**The user quickly rejects false positives among these potential instances in a quick manual examination of the full list.**

# Use of SREE, Cont'd

**Any ambiguity whose finding requires**

- **parsing of NL sentences,**
- **correct part-of-speech identification,**
- **seeing context, or**
- **understanding semantics**

**is left for manual searching.**

# SREE's Design Rationale

**SREE has 100% recall for the ambiguities in its clearly specified domain, ...**

**but less than 100% *precision* for these same ambiguities, ...**

**since it finds, e.g., all instances of “only”, not just the ambiguous ones.**



# SREE's Design, Con'd

**The analyst can quickly eliminate the false positives in SREE's output**

**and then focus attention on the ambiguities that are outside SREE's clearly specified domain.**

# Enhancement of Dekhtyar & al

**Humans vetting the poorer of two tools did a better job, as if they sensed the poor quality and rose to the occasion.**

**So maybe take the best tool available and randomly split its output to two groups of vetters.**

**BOBW!**

# Future Research Agenda

**For each RE task to which NLP tools are being applied, e.g.,**

- **abstraction identification,**
- **ambiguity identification, and**
- **tracing,**

# Future Research Agenda, Cont'd

try to find an *algorithmically identifiable* partition of the task into

1. a *clerical* part that can be done by a dumb tool with 100% recall and not too much imprecision and
2. a *thinking-required* part that must be left to a human analyst to do manually.

# Research Required

**Finding this partition for any task will require research to think of a different way to decompose the task.**

**It will require a thorough understanding of the task and of what is algorithmically possible.**

# Research Required, Cont'd

**For any task, the partitioning will take into account**

- **the burden to the human analyst of the imprecision of the clerical part and**
- **the difficulty to the human analyst of the thinking-required part.**

# Research Required, Cont'd

**Obtaining this information will require research like that done by Dekhtyar *et al.* for tracing tools to determine**

- **what is really difficult for humans and**
- **how well humans perform parts of the task with and without automation.**

# Read Our Paper

**Now go read our paper!**

**Write a rebuttal!**

**Join in on the research!**

**But, please be polite and stay for the rest of the talks of this session!**

