

Автоматический перевод: достижения и проблемы

Выполнил: Слушатель

Гусев Д.А.

Основные понятия

Машинный перевод — процесс перевода текстов с одного естественного языка на другой полностью специальной компьютерной программой.

Формы взаимодействия ЭВМ и человека:

С постредактированием: исходный текст перерабатывается машиной, а человек-редактор исправляет результат.

С предредактированием: человек приспособливает текст к обработке машиной (устраняет возможные неоднозначные прочтения, упрощает и размечает текст), после чего начинается программная обработка.

С интерредактированием: человек вмешивается в работу системы перевода, разрешая трудные случаи.

Смешанные системы (например, одновременно с пред- и постредактированием).

Автоматизированный перевод - перевод текстов на компьютере с использованием компьютерных технологий - программа просто *помогает* человеку переводить тексты.

Формы взаимодействия ЭВМ и человека:

Частично автоматизированный перевод: например, использование переводчиком-человеком компьютерных словарей.

Системы с разделением труда: компьютер обучен переводить только фразы жёстко заданной структуры (но делает это так, чтобы исправлять за ним не требовалось), а всё не уложившееся в схему отдаёт человеку.

История становления отрасли

1946 год – первое упоминание о идее применения ЭВМ при переводе

1954 год - Джорджтаунский эксперимент

1960-е годы – появление двух систем русско-английского перевода

MARK (в Департаменте иностранной техники ВВС США);

GAT (разработка Джорджтаунского университета)

1980-е годы - сложился рынок коммерческих разработок переводческих систем

Основные виды машинного перевода

1. Перевод, основанный на правилах – базируется на определении связи структуры входного и выходного предложения.

прямой перевод

трансфер

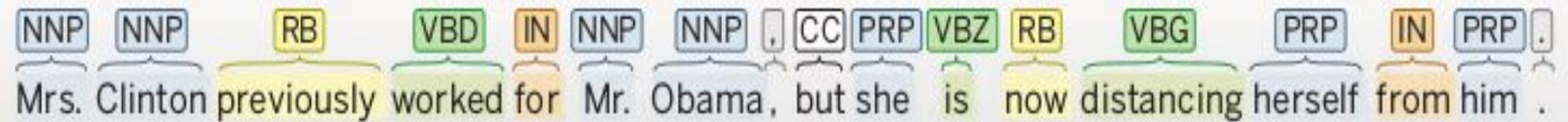
интерлингва

2. Перевод, основанный на статистике -обучение машины посредством предоставления достаточно большого (сотни тысяч) количества параллельных текстов — содержащих одинаковую информацию на разных языках.

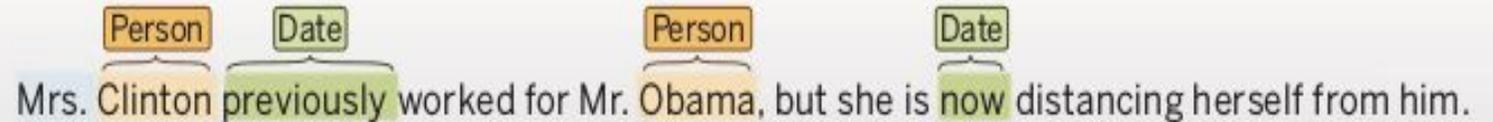
Машинный перевод на базе фраз — это самая простая и популярная версия статистического машинного перевода

3. Нейромашинный перевод

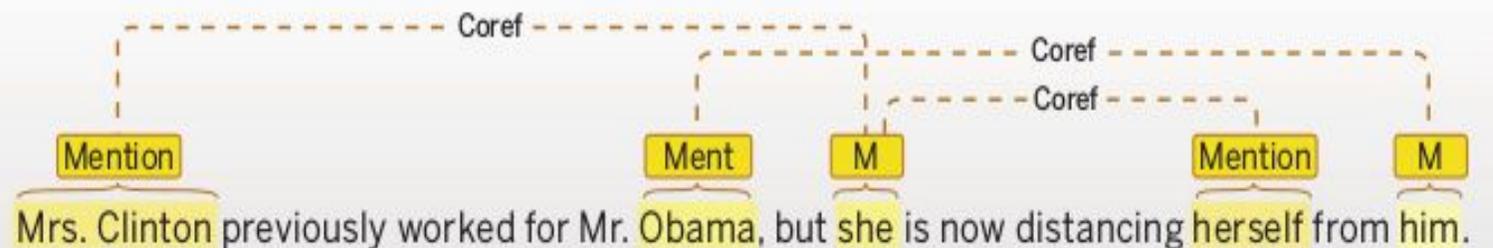
Part of speech:



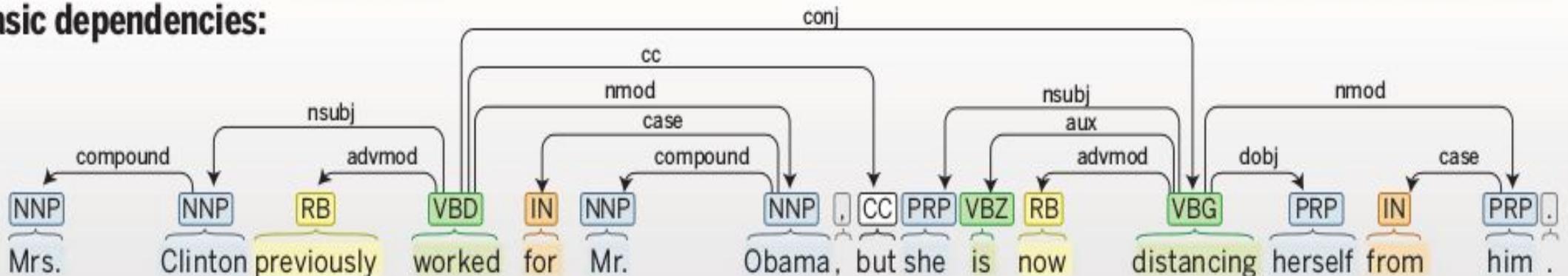
Named entity recognition:



Co-reference:



Basic dependencies:



Интерлингвистические системы

В основе интерлингвистических систем лежит идея существования универсального метаязыка, представляющего смысл предложения на любом естественном языке. Такие системы включают в себя два этапа: анализ и синтез. На этапе анализа входной текст трансформируется при помощи словаря и грамматических правил исходного языка в представление на универсальном метаязыке. На втором этапе это представление преобразуется в предложение выходного языка при помощи словаря и грамматических правил языка перевода.

Недостатки и преимущества систем, основанных на правилах

Основной недостаток - неразрешенная проблема нахождения универсального для всех естественных языков смыслового представления.

Основное преимущество - высокая точность перевода. Однако, вместе с ней нередко появляется некоторый “машинный” акцент, неестественность выходного текста

Статистический машинный перевод

Статистический машинный перевод основан на поиске наиболее вероятного перевода предложения, с использованием данных двуязычного корпуса (Parallel Corpora) Битекст. В результате при выполнении перевода компьютер не оперирует лингвистическими алгоритмами, а вычисляет вероятность применения того или иного слова или выражения. Слово или последовательность слов, имеющие оптимальную вероятность, считаются наиболее соответствующими переводу исходного текста и подставляются компьютером в получаемый в результате текст. В статистическом машинном переводе ставится задача не перевода текста, а задача его расшифровки.

Now I shall tell you a story which will show you how honest I have always been all my life.

"Did you live in Washington in 1867?" the general asked me.

"Yes, I did," I answered.

"Are you looking for a dog, sir?" I asked.

"Oh, yes! Have you seen it?" said the man.

"Your dog was here a few minutes ago and I saw how it went away with a man," I said. "If you want, I shall try to find it for you."

Теперь я расскажу вам историю, которая покажет вам, насколько я честен, что всегда был на всю жизнь.

«Вы жили в Вашингтоне в 1867 году?» - спросил меня генерал.

«Да, я сделал», ответил я.

«Вы ищете собаку, сэр?» Я спросил.

«О, да, вы видели это?» - сказал мужчина.

«Ваша собака была здесь несколько минут назад, и я увидел, как она ушла с мужчиной», - сказал я. «Если хочешь, я постараюсь найти его для тебя».

Анализ «состояния рассказчика»

«Завтрак на Бродвее» — новое место, которое специализируется, как вы могли догадаться, на завтраках. Был там прошлым воскресеньем в районе часа. **Еда была неплохой**, но обслуживание — достаточно ужасное. Прошло 15 минут, пока нам принесли меню, и ещё 30 мы ждали еду. И клиентов особо не было — было занято всего несколько столов! **Если вы не против ожидания**, то цена соответствуют качеству. Я, наверно, попробую зайти в другой раз. Возможно, им нужно время, чтобы наладить работу.

«Завтрак на Бродвее» — новое место, которое специализируется, как вы могли догадаться, на завтраках. Был там прошлым воскресеньем в районе часа. Еда была неплохой, но **[гнев: обслуживание — достаточно ужасное]**. **[Отвращение: Прошло 15 минут, пока нам принесли меню, и ещё 30 мы ждали еду. И клиентов особо не было — было занято всего несколько столов!]** Если вы не против ожидания, то цена соответствуют качеству. Я, наверно, попробую зайти в другой раз. **[Неопределённость: Возможно, им нужно время, чтобы наладить работу]**.