



das Nutella oder die Nutella?

Grammatical Gender Prediction of German Nouns

(UseR! 2019, Toulouse, France)

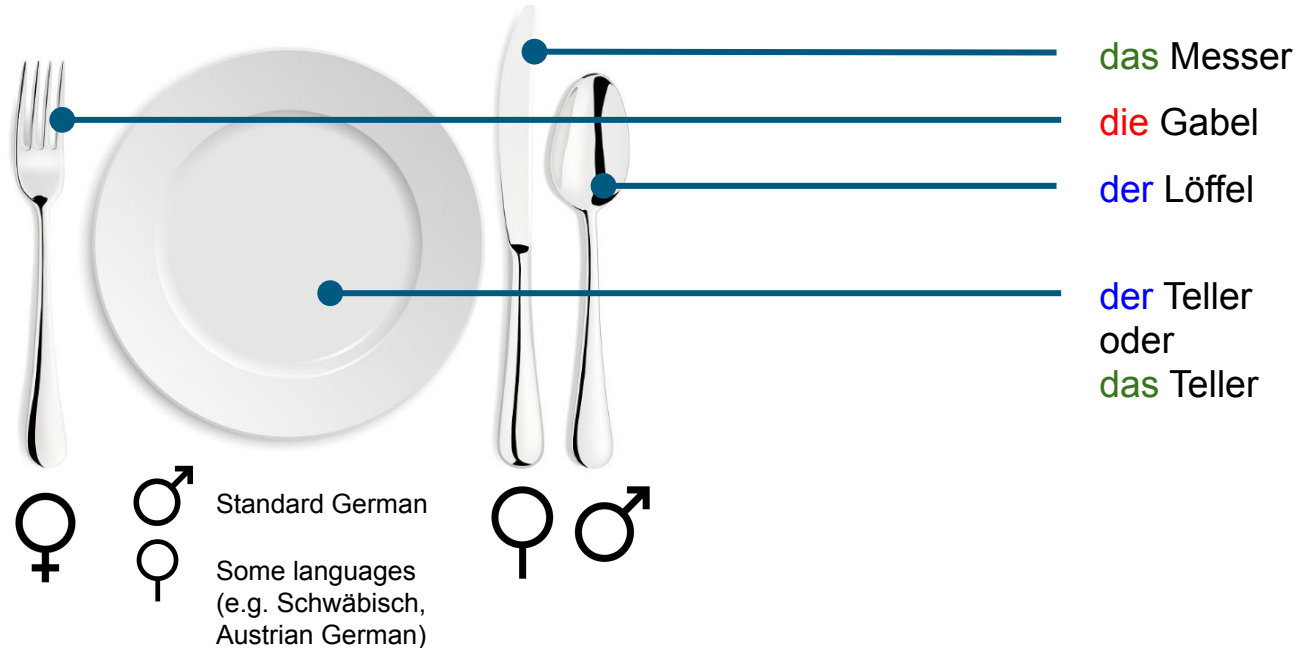
Chung-hong Chan, PhD

I'll tell you

- how to train a machine learning algorithm with a very high accuracy metric, but miserably fails the intended task.
- What are *Money Cats*?

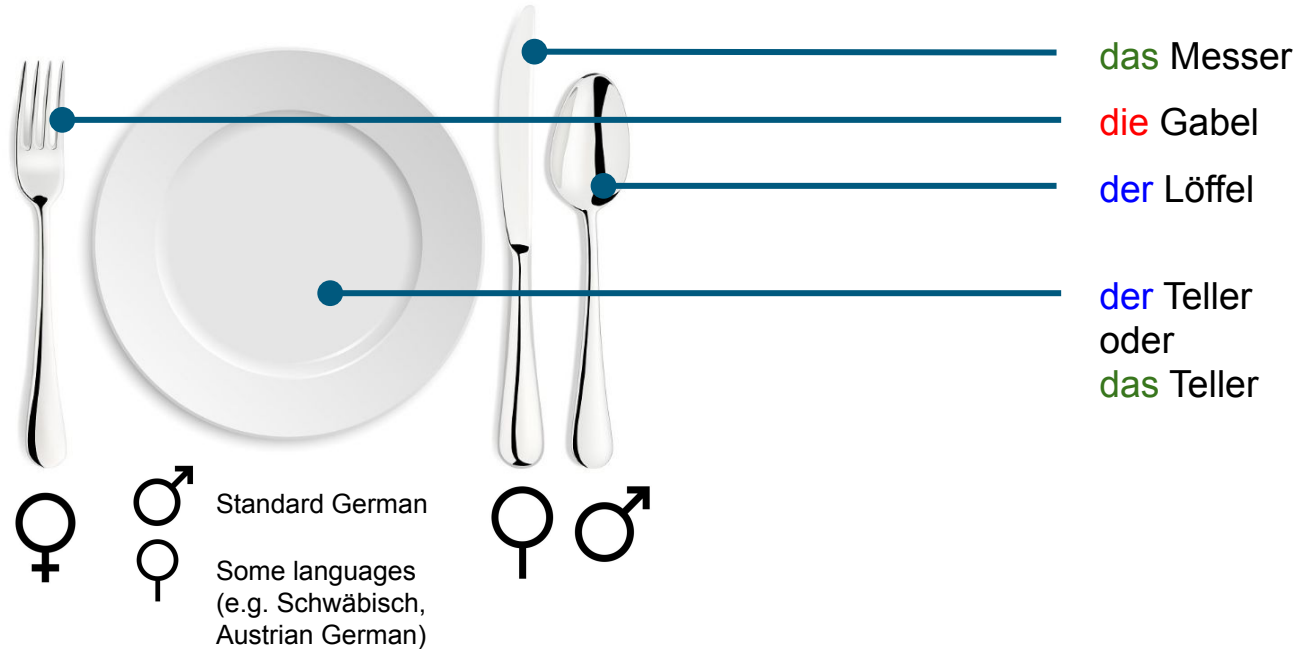
What is grammatical gender?

English	German	French
I have a fork.	Ich habe eine Gabel.	J'ai une fourchette.
I have a spoon.	Ich habe einen Löffel.	J'ai une cuillère.
I have a knife.	Ich habe ein Messer.	J'ai un couteau.



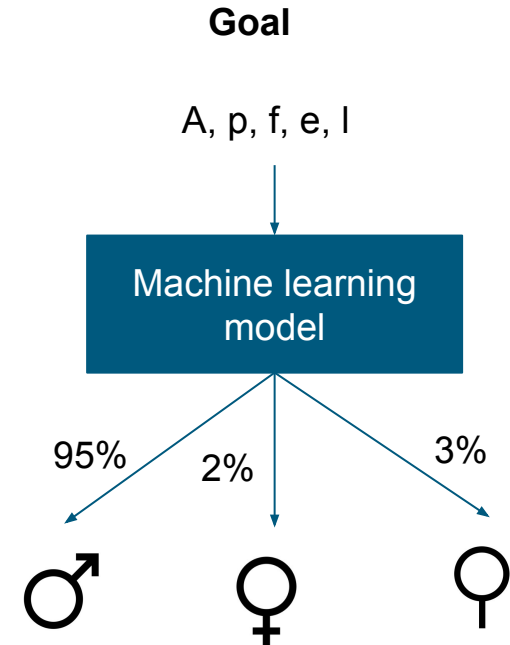
What is grammatical gender?

English	German	French
I have a fork.	Ich habe eine Gabel.	J'ai une fourchette.
I have a spoon.	Ich habe einen Löffel.	J'ai une cuillère.
I have a knife.	Ich habe ein Messer.	J'ai un couteau.



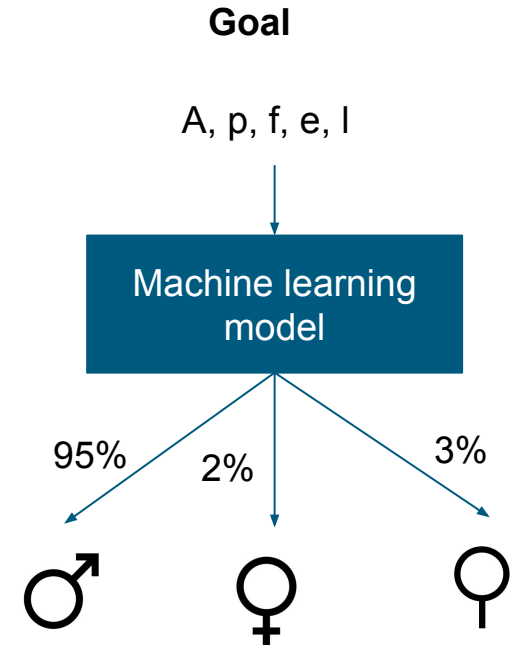
Is it completely random?

die Gesundheit der Vater
die Einsamkeit die Schwester
die Krankheit das Fenster
der Terrorismus
der Kapitalismus
das Mädchen
das Häuschen



Is it completely random?

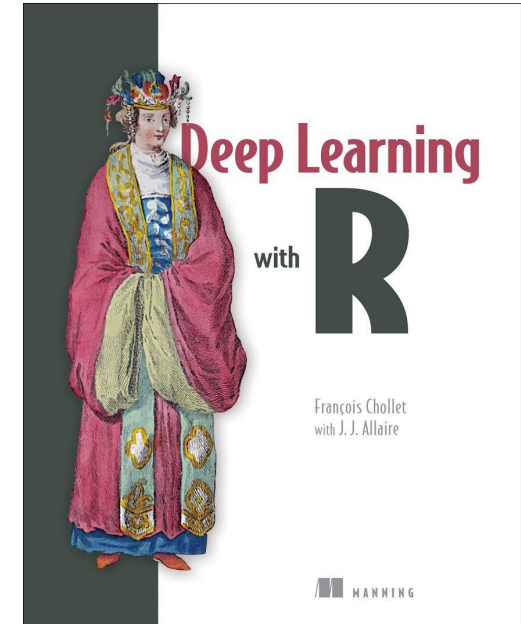
die Gesundheit der Vater
die Einsamkeit die Schwester
die Krankheit das Fenster
der Terrorismus
der Kapitalismus
das Mädchen
das Häuschen



Model

- 329,876 nouns with their respective gender
- LSTM model

```
model %>%  
  layer_embedding(input_dim = 66, output_dim =  
128) %>%  
  layer_lstm(units = 96, dropout = 0.05,  
recurrent_dropout = 0.05, return_sequence = TRUE) %>%  
  layer_lstm(units = 64, dropout = 0.05,  
recurrent_dropout = 0.05) %>%  
  layer_dense(units = 3, activation = 'sigmoid') %>%  
  compile(loss = 'binary_crossentropy', optimizer =  
'adam', metrics = c('accuracy')) %>%  
  fit(X, y, batch_size = 300, epochs = 30,  
validation_split = 0.20)
```



Results

- Accuracy: 97.8%
 - *Gesundheit* / Health - 99.6% ♀ (Correct)
 - *Zeitung* / Newspaper - 99.78% ♀ (Correct)
 - *Mädchen* / Girl - 99.36% ♀ (Correct)
 - *Terrorismus* / Terrorism - 99.72% ♂ (Correct)

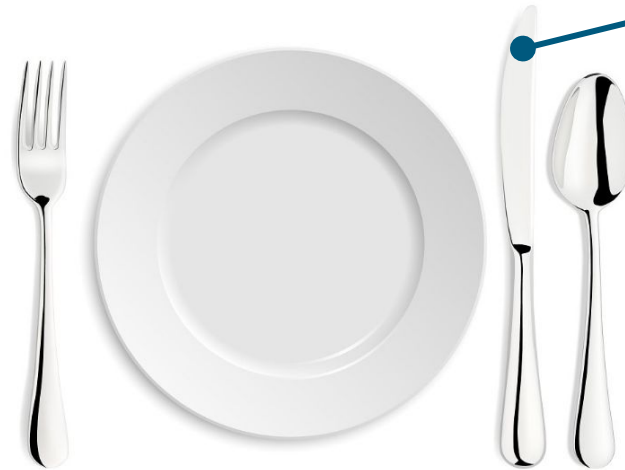
Results (Common Nouns)



68.42% ♂
19.46% ♀
(Wrong!)

- *Schüssel* / Bowl - 81.03% ♂
- *Name* / Name - 44.01% ♀
- *Käse* / Cheese - 73.43% ♀

Results (Common Nouns)



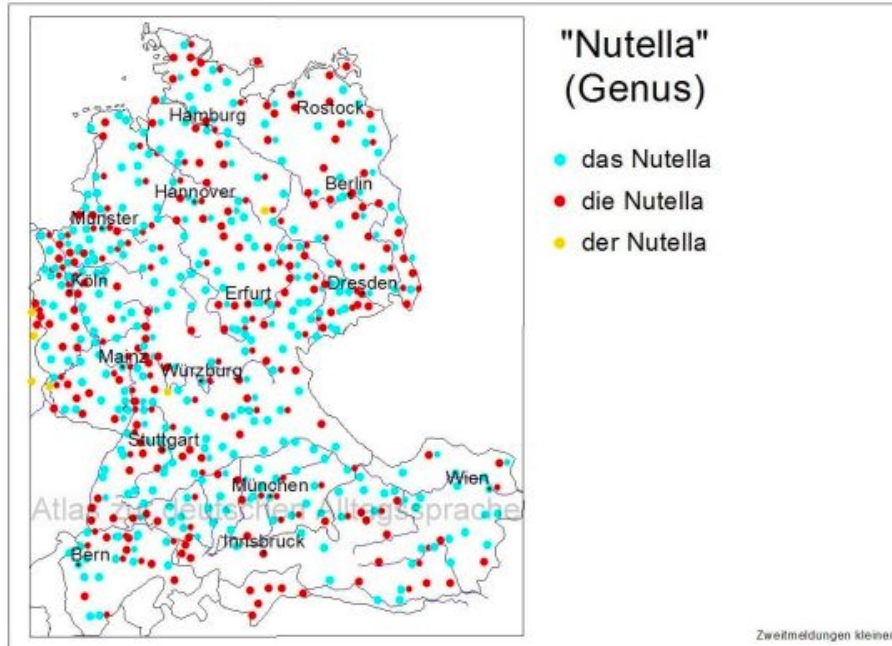
68.42% ♂
19.46% ♀
(Wrong!)

- *Schüssel* / Bowl - 81.03% ♂
- *Name* / Name - 44.01% ♀
- *Käse* / Cheese - 73.43% ♀

Because

*Schweinefleischetikettierung
überwachungs
aufgabenübertragungs
gesetz*

das Nutella or die Nutella?



♀ 51.59%

Quelle: Atlas zur deutschen Alltagssprache, Universität Augsburg

DeepDream

generate random German nouns

*Selbstwarenken, Sexualandel, Mitter, Gestand, Gesamik,
Herzstrationsschlag, Hauptster, Panapie, Melonessel,
Kunstlingsstellung, Transchaftsschwein, Parapfen, Einzelle, Flecke,
Finante, Schleppenstellung, Hodismus, Amente, Bestabilität,
Haustrieb, Werkschen, Grenzeichskenraum, Erdbeinkontrolle,
Nachtbaum, Siebel, Stacherung, Polyché, Erganismus,
Einspeziehung, Homologe, Laufwand, Blauchmelze, Erdausgabe,
Straßenarbe, Fernschaft, Konzereiter, Abrief, Halbbaum, Kreise,
Gelbst, Haushaltssport, Hausel, Stache, Rotbaum, Kommung,
Schwanzahl, Freihe, Geldkatzen*

DeepDream

generate random German nouns

Selbstwarenken, Sexualandel, Mitter, Gestand, Gesamik, Herzstrationsschlag, Hauptster, Panapie, Melonessel, Kunstlingsstellung, Transchaftsschwein, Parapfen, Einzelle, Flecke, Finante, Schleppenstellung, Hodismus, Amente, Bestabilität, Haustrieb, Werkschen, Grenzeichskenraum, Erdbeinkontrolle, Nachtbaum, Siebel, Stacherung, Polychy, Erganismus, Einspeziehung, Homologe, Laufwand, Blauchmelze, Erdausgabe, Straßenaarbe, Fernschaft, Konzereiter, Abrief, Halbbaum, Kreise, Gelbst, Haushaltspport, Hausel, Stache, Rotbaum, Kommung, Schwanzzahl, Freihe, Geldkatzen



“Money Cats”



多謝晒

Thank you very much

Danke schön

どうもありがとうございました

대단히 감사합니다.