

Ausschreibung einer Studien/Diplomarbeit zum Thema

”Lesbarkeit”

August 18, 2010

Betreuer: apl. Prof. Dr. Uwe Reyle

Worum geht es?

Für das Verständnis von Texten sind nicht nur die Bildungsvoraussetzungen der Leser entscheidend. Textverständnis ergibt sich auch und wesentlich aus dem Zusammenspiel folgender Merkmale und Sachverhalte.

1. Komplexität des Gegenstandsbereiches und begriffliche Schwierigkeiten
2. sprachliche Komplexität auf den Ebenen der Wortbildung, der Satzkonstruktion, der Satzinhalte sowie der inhaltlichen Beziehungen zwischen Sätzen
3. Qualität des Stils, Leserlichkeit und Gestaltung des Druckbildes, Leserbezug

Maße für die Lesbarkeit oder den Schwierigkeitsgrad von Texten (engl.: readability-measure) vernachlässigen die unter 1. und 3. genannten Parameter vollständig und basieren bzgl. der Parameter unter 2. ausnahmslos auf Oberflächenmerkmalen. Messbarkeit beruht hier auf der Annahme, dass die Lesbarkeit eines Textes mit den Durchschnittswerten einer Reihe von einfach zugänglichen Sprachfaktoren korreliert. Die wichtigsten solcher Faktoren sind:

1. SL = durchschnittliche Satzlänge in Wörtern
2. WL = durchschnittliche Wortlänge in Silben
3. ES = Prozentsatz der einsilbigen Wörter
4. MS = Prozentsatz der drei- und mehrsilbigen Wörter
5. SW = Prozentsatz der seltenen Wörter, d.h. der Wörter, die nicht in der Liste der 1000 häufigsten Wörter vorkommen
6. IW = Prozentsatz der langen Wörter, d.h. der Wörter mit mehr als sechs Buchstaben

Als Korrelation zwischen Worteigenschaften und Schulstufen gibt Malewski¹ folgende Werte an.

Schulstufe	SL	WL	ES	MS	sW	IW	Lix
1	7	1,43	6,2	5	10	12	19
2	9	1,51	57	7	11	15	24
3	10	1,55	55	10	14	17	27
4	11	1,58	53	11	17	19	30
5	12	1,61	52	12	19	20	32
6	13	1,65	51	14	20	21	34
7	14	1,67	50	15	22	22	36
8	15	1,70	49	16	23	23	38

¹In: tekompnachrichten | 19.Jahrgang | 3 / 1997 | Seite 30

Was ist zu tun

Da die meisten Untersuchungen zur Lesbarkeit in den USA, also für das Englische durchgeführt wurden, sollen zunächst anhand geeigneter Korpora entsprechende Vergleichsdaten für das Deutsche erfasst werden. Die Korpora sollen zumindest aus den folgenden Bereichen stammen bzw. erstellt werden:

1. Wissenschaftliche Texte
2. Bild Zeitung (Verfügbarkeit muss noch geklärt werden)
3. Stuttgarter Zeitung
4. Kinderliteratur

In einem zweiten Teil sollen die am IMS vorhandenen Ressourcen eingesetzt werden, um die Ergebnisse zu verbessern. Das Maß WL könnte man beispielsweise mit "durchschnittlicher Wortlänge in Morphemen" ersetzen oder gar durch ein selbst entwickeltes Maß, das eine linguistisch fundiertere Berechnung der Komplexität (etwa mittels derivationsmorphologischer und kompositionsmorphologischer Merkmale) erlaubt. Ferner können auch auf Satzebene strukturelle Merkmale herangezogen werden (abgeleitet aus Chunks oder Parse-Bäumen), mit denen die Komplexität des Satzaufbaus bestimmt werden kann.

Die Entwicklung dieser komplexeren Messverfahren ist jedoch eher der Gegenstand einer Diplomarbeit. Dies wird im Rahmen einer Studienarbeit nur marginalen Charakter haben.