

Terminologieextraktion mit Acrolinx IQ

–

Verbesserung von Unternehmenssprache und Übersetzungsprozess

Melanie Siegel

Einführung

In jedem Unternehmen kommt es früher oder später zu dem Punkt, an dem es notwendig wird, sich mit Terminologie zu beschäftigen. Schon für die interne Kommunikation im Unternehmen ist es unerlässlich, dass die Mitarbeiter dieselbe Sprache sprechen und Produkte und Prozesse so benennen, dass es keine Missverständnisse gibt. Eine inkonsistente Kommunikation nach außen führt zu Verwirrung bei Kunden und potentiellen Kunden und beeinträchtigt den Unternehmenserfolg erheblich.

Der Deutsche Terminologie-Tag (DTT) beschreibt die Prozesse der Terminologiearbeit in seiner Schrift „Best-Practices-Ordner: Terminologiearbeit“ (<http://www.iim.fh-koeln.de/dtt/>)

Der erste Schritt für das Unternehmen ist zu identifizieren, was Terminologie für dieses Unternehmen eigentlich bedeutet. Das Unternehmen muss definieren, welche Terme relevant sind und welche Benennungen verwendet werden sollen. Idealerweise sollte auch eine Liste der nicht zu verwendenden Benennungen erstellt werden, damit die Konsistenz gewährleistet ist.



Abbildung 1: Definitionen für Terminologie mit wordle.net

Das „Semiotische Dreieck“ von Ogden & Richards (1923) zeigt, wie Begriffe, Benennungen und Objekte der Welt miteinander in Beziehung stehen.

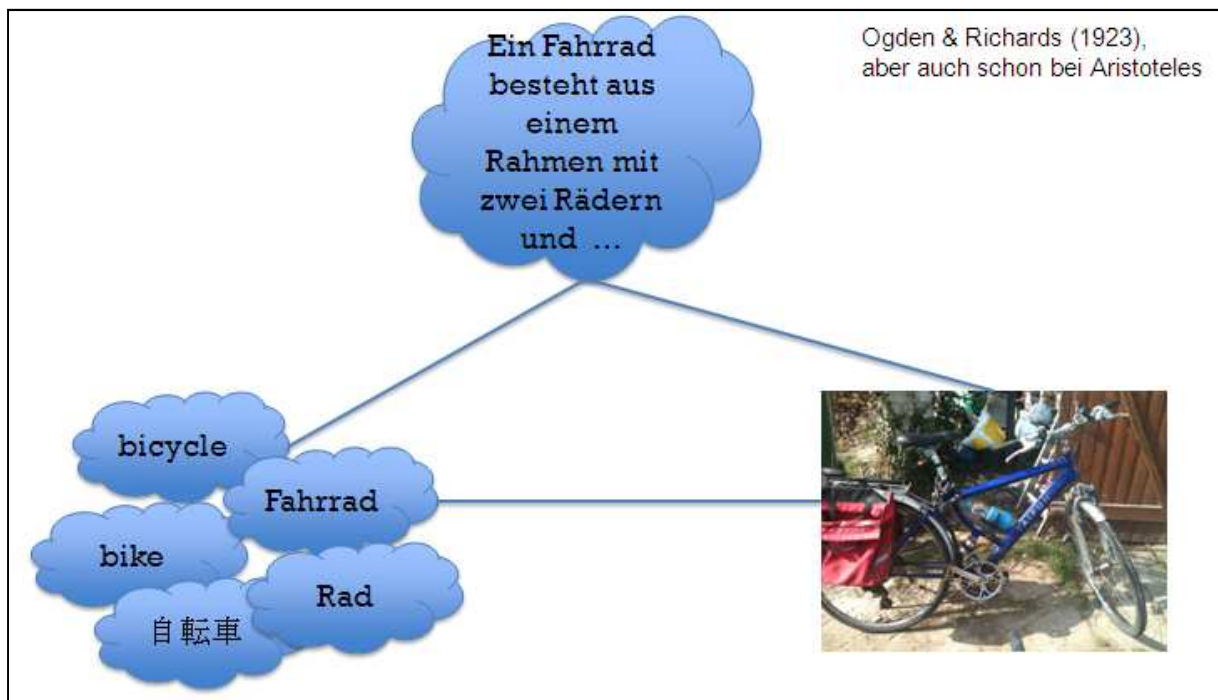


Abbildung 2: Semiotisches Dreieck

Aber wofür dient Terminologie im Unternehmen?

Es ist notwendig, die Motivation für die Terminologiearbeit im Unternehmen klar herauszuarbeiten, denn schließlich kann Terminologiearbeit durchaus aufwändig sein. Wir geben hier einige Hinweise aus Erfahrungen in sehr unterschiedlichen Unternehmen.

1. Das **Firmenprofil** wird durch den Gebrauch von Terminologie bestimmt. Kunden und potentielle Kunden werden verwirrt, wenn dieselben Produkte in Broschüren, Handbüchern und auf Internetseiten unterschiedliche Benennungen haben. Konsistente Terminologie in der Außenwirkung des Unternehmens erhöht die Kundenbindung und die Nutzbarkeit von Produkten.
2. Die **Suche im Internet** nach Produkten des Unternehmens wird gestärkt, wenn die Terminologie konsistent ist. Man spricht hier von *Search Engine Optimization*.
3. Der **Support** wird erheblich vereinfacht, wenn die Kunden in ihren Anfragen dieselben Benennungen verwenden, weil diese auch in allen Produktinformationen verwendet werden.
4. Der **Übersetzungsprozess** wird einfacher, konsistenter und letztlich auch preisgünstiger, wenn die Terminologie in den Texten konsistent verwendet wird. Im Rahmen der Globalisierung müssen viele Unternehmen die Dokumentation Ihrer Produkte in mehrere Sprachen übersetzen. Auch die Internet-Seiten sind häufig mehrsprachig, um Kunden in anderen Ländern zu erreichen. Wenn die Terminologie im Ausgangstext konsistent ist, werden Satzvarianten reduziert und die Übersetzung wird einfacher, schneller und günstiger. Wenn dazu kommt, dass auch die Terminologie für die Zielsprachen konsistent festgelegt und prüfbar ist, dann werden die Dokumentation und die Präsentation der Dokumente über die Sprachen hinweg konsistent.
5. Eine konsistente Terminologie unterstützt den **Schreibprozess** erheblich, wenn sie mit einer Terminologieprüfung gekoppelt ist. Die Autoren und Autorinnen werden in ihrem Bemühen, konsistent und verständlich zu schreiben, effektiv unterstützt. Das Korrekturlesen wird erheblich vereinfacht.
6. Der **Kommunikationsprozess innerhalb des Unternehmens** wird gestützt. Im Unternehmen sprechen alle Mitarbeiter dieselbe Sprache. Es ist nicht mehr notwendig, in Meetings über Begriffsdefinitionen zu streiten, wenn diese bereits festgelegt sind. Neue Kollegen und Kolleginnen verstehen die Produkte und Prozesse schneller, wenn alle Produkte und Prozesse mit denselben Benennungen versehen werden.

Woher kommt die Terminologie?

Es ist in den seltensten Fällen so, dass ein Unternehmen vor der Entwicklung der ersten Produkte die gesamte Terminologie aufgestellt hat. Vielmehr gibt es ein Sammelsurium:

- Angefangene Terminologiedatenbanken, die in verschiedenen Formaten vorliegen.
- Wortlisten, die von Autoren und Autorinnen im Schreibprozess aufgestellt worden sind.
- Mehrsprachige Wortlisten der Übersetzer, die häufig als Excel-Listen vorliegen.
- Produktlisten, in denen die verschiedenen Produkte und ihre Produktvarianten aufgelistet sind.

Wesentlich informativer als alte Listen sind jedoch die Textdokumente, die bereits geschrieben wurden, da es sich hier häufig um große Textmengen handelt. Diese Texte haben den großen Vorteil, dass dort die tatsächlich verwendete Terminologie zu finden ist. Wenn eine Terminologiedatenbank aufgesetzt wird, die unabhängig von den bereits geschriebenen Texten ist, dann kann man nicht sicher sein, dass diese Datenbank die

Benennungen enthält, die auch wirklich gebraucht werden. Daher ist es wichtig, mit einem Terminologieextraktionsprozess auf diese Texte zuzugreifen.

Seit einigen Jahren werden in den Unternehmen für den Übersetzungsprozess Translation Memories eingesetzt. Auch diese Translation Memories sind eine wichtige Wissensquelle für die multilinguale Terminologie, denn hier sind die Übersetzungen für die Benennungen bereits enthalten. Selbst wenn die Terminologie für das Unternehmen perfekt ist, muss gewährleistet sein, dass die Übersetzer sie auch konsistent einsetzen.

Wie funktioniert die Terminologieextraktion mit Acrolinx IQ?

Die Extraktion von Terminologie aus Textdokumenten basiert auf linguistischen Informationen. Die Texte werden zunächst mit linguistischen Informationen angereichert und anschließend werden Benennungen gezielt extrahiert, zum Beispiel Nomen, Komposita, Akronyme und Produktnamen. Zusätzliche Informationen werden bei der Extraktion berücksichtigt, wie z.B. Informationen über mögliche Rechtschreibfehler, unbekannte Wörter, generische Wörter der Sprache, die Häufigkeit des Wortes in den Texten oder Wörter, die schon in der Terminologiedatenbank enthalten sind. Die Extraktion gibt den Anwendern Informationen, die für den Validierungsprozess wichtig sind: Die Häufigkeit des Auftretens eines Wortes in den Texten, Satzkontexte und Termvarianten, die in den Texten auftreten.

Term	Frequency	Status	Approved Terms	Contexts	Files
Nip	4	Proposed		Rotierende Bauteile Hantieren mit Werkzeugen an Bauteilen während des Betriebs Überqueren der Maschine an nicht geeigneten Bauteilen	
Bauteil	3	Proposed		Hydraulik An hydraulischen Einrichtungen darf nur Personal mit speziellen Kenntnissen und Erfahrungen in der Hydraulik arbeiten.	
Hydraulik	3	Proposed		Hydraulik Notfallvorrichtungen Hydraulik, Pneumatik Lärm Gefahrgutzeichen / Schilder Notfallvorrichtungen Öle, Fette, andere chemische Substanzen Besondere Gefährdungsstellen Häufige Unfallursachen	
Notfallvorrichtung	3	Proposed		Notfallvorrichtungen Hydraulik, Pneumatik Lärm Gefahrgutzeichen / Schilder Notfallvorrichtungen Öle, Fette, andere chemische Substanzen Besondere Gefährdungsstellen Häufige Unfallursachen	
Unfallursache	3	Proposed		Häufige Unfallursachen	
Bespannung	2	Proposed		Aufaufstellen zwischen rotierenden Walzen und Bespannungen Berühren von Papierbahn, Bespannungen oder Seilen	
Entwässerung	2	Proposed		- Verbesserte Entwässerung nur durch Erhöhen der Verweilzeit im Nip möglich. - Verbesserte Entwässerung durch erhöhten Flächendruck möglich.	
Explosionsgefahr	2	Proposed		Z. B. kann Brand- und Explosionsgefahr bestehen. Vor dem Schweißen, Brennen und Schleifen Maschine und deren Umgebung von Staub und brennbaren Stoffen reinigen und für ausreichende Lüftung sorgen (Explosionsgefahr). Hydraulisch begrenzter Nip - Hydraulischer Widerstand im Faservlies überwiegt. Strukturbegrenzter Nip - Strukturwiderstand im Faservlies überwiegt	
Faservlies	2	Proposed			

Abbildung 3: Ergebnisse des Termextraktionsprozesses (monolingual)

Die extrahierten Benennungen können so von den Terminologieexperten validiert und freigegeben werden. Damit ist es schon möglich, die Ausgangssprache zu prüfen und Konsistenz herzustellen.

Aber wir gehen noch einen Schritt weiter: Wir wollen Terminologie in mehreren Sprachen extrahieren.

Wie funktioniert die Multilinguale Termextraktion mit Acrolinx IQ?

Die multilinguale Termextraktion basiert auf der allgemeinen Termextraktion mit linguistischen Mitteln, wie sie oben beschrieben ist. Im Gegensatz dazu verwenden wir hier aber ein statistisches Verfahren, das auch mit weniger linguistischen Informationen über die Zielsprache möglich ist. Wenn eine der beteiligten Sprachen eine Acrolinx bekannte Sprache – Englisch, Deutsch, Französisch, Schwedisch oder Japanisch – ist, dann können die Zielsprachen beliebig sein. Es ist möglich, Benennungen in zwei, drei oder auch mehr Sprachen gleichzeitig zu extrahieren.

Für die multilinguale Termextraktion benötigen wir Daten, bei denen Sätze in den zu verwendenden Sprachen miteinander verbunden (aligniert) sind. Daher werden hierfür häufig Translation Memory-Dateien im TMX-Format verwendet, wie sie in vielen Unternehmen vorliegen. Prinzipiell ist es auch möglich, mit parallelen Textdateien zu arbeiten, bei denen die Sätze in den jeweiligen Sprachen einander zugeordnet sind. Daneben kann man auch mit parallelen, übersetzten Dokumenten arbeiten und eine automatische Alignierung der Sätze vorschalten. Allerdings müssen diese Dokumente sehr ähnliche Textstrukturen aufweisen, damit die übersetzten Sätze auch gefunden werden können.

Die gesamte linguistische Technologie analysiert die Sätze in der bekannten Sprache und extrahiert Terminologie. Dann wird die so extrahierte Terminologie mit den Benennungen der Zielsprache in Beziehung gesetzt. Dafür nutzen wir statistische Verfahren und zusätzliche linguistische Informationen, je nachdem wie viele Informationen über diese Sprache zur Verfügung stehen.

Das Ergebnis ist eine Excel-Tabelle mit Benennungen, Übersetzungen, Frequenzinformationen und Informationen über die Satzkontexte.

1	Term Candidate (de)	Frequency	Status	Language	Proposed Translation	Frequency	Contexts
16	Klappenantrieb	56					
17				fr	servomoteur de volet	17	Stellsignal Klappenantrieb ↔ le de volet
18				fr	servomoteur volet	4	2 Klappenantrieb ↔ 2 Servomot Klappenantrieb 2 - 6.00% ↔ 2
19				fr	servomoteur à volet	4	Bei Soll- Istwertabweichung wi der Volumenstrom über der Mo
20				fr	servomoteur de volets	3	Klappenantrieb (SUT) mit LON Modul ↔ Servomoteur de volet
21				en	damper drive	31	Klappenantrieb ↔ Damper driv
22				en	valve drive	3	Klappenantrieb (SUT) mit LON drive / SUT) with LON commu
23				en	damper size	2	Integrierter Klappenantrieb (Fo continued)
24				en	damper actuator	1	Durch Schalten der Kontakte wi Volumenstromer ein neuer
26	Stellantrieb	37					
27				fr	servomoteur	9	für Stellantrieb ↔ pour servom
28				fr	servomot.	1	für Stellantrieb / Auf/Stop/Zu)
29				en	actuator	16	für Stellantrieb ↔ for actuator
30							Die Pumpe geht in Betrieb - sch
31	Ventil	1095					

Abbildung 4: Beispiel für Ergebnisse der Multilingualen Termextraktion

Diese Ergebnistabelle liefert interessante Informationen über die Inhalte der Dokumente in allen beteiligten Sprachen. Einige Beispiele für Benennungen, die in Deutsch-Englischen Translation Memories gefunden wurden:

- Hubtisch
 - lift table
 - table lift
 - lifting platform
- Grundstellung
 - starting position
 - basic position
 - home position
 - basic setting
 - initial position
 - starting pos.
 - normal position
- adaptor
 - Zwischenstück
 - Adaptor
- operating unit

- Bedieneinheit
- Bediengerät

Die Extraktion enthält Informationen über Übersetzungsvarianten, aber auch über Varianten und Synonyme in allen beteiligten Sprachen.

Selbstverständlich ist auch hier eine Validierung durch Experten notwendig. Die Software kann nicht entscheiden, ob „operating unit“ z.B. durch „Bedieneinheit“ oder „Bediengerät“ übersetzt werden soll. Aber das Extraktionsergebnis gibt wertvolle Hinweise.

Was kann man mit extrahierter Terminologie machen?

Zunächst gibt die extrahierte multilinguale Terminologie wichtige Hinweise für die Qualität der Texte in den beteiligten Sprachen und für die Qualität der Übersetzungen. Mit diesen Informationen lassen sich Texte und Translation Memories nachhaltig verbessern. Für die Übersetzer kann nun ein mehrsprachiges Wörterbuch für die im Unternehmen relevanten Benennungen zur Verfügung gestellt werden. Eine Prüfung auf Terminologiekonsistenz mit Acrolinx IQ ist dann auch für weitere Sprachen möglich. Die Übersetzer können damit die Qualität ihrer Übersetzungen erhöhen, indem sie Benennungen nachschlagen und ihre Texte automatisch auf konsistente Terminologie prüfen.

Da auch Varianten und Synonyme im Ausgangstext gefunden werden, haben die Autoren eine effiziente Methode, die terminologische Konsistenz ihrer Ausgangstexte zu gewährleisten.

Schließlich können gefundene mehrsprachige Synonyme auch dafür genutzt werden, die Suche im Internet zu verbessern, indem diese Synonyme konsequent in Webseiten eingepflegt werden.

Zusammenfassung

Der Aufbau von Terminologie ist eine wichtige Grundlage für die konsistente Kommunikation innerhalb und außerhalb des Unternehmens. Terminologearbeit ist notwendig für die Außendarstellung des Unternehmens, den Workflow und für alle Arten interner und externer Kommunikation. Dabei müssen Benennungen berücksichtigt werden, die im Unternehmen wirklich relevant sind.

Um das zu erreichen, ist für den Aufbau von Terminologie die Extraktion von Benennungen aus bereits geschriebenen Texten der richtige Weg. Die multilinguale Termextraktion von Acrolinx IQ bietet darüber hinaus die Möglichkeit zur Extraktion von mehrsprachigen Benennungen auf der Basis linguistischer Intelligenz.