

# Sprachpolitik im Internet

Hausarbeit im Rahmen des Hauptseminars Sprachpolitik von Frau Prof. Dr. Eschbach-Szabo im Sommersemester 2005 am Seminar Japanologie der Fakultät für Kulturwissenschaften der Universität Tübingen

Datum: 27.10.2005  
Adresse: Markus Stengel  
Herrenberger Str. 22  
72070 Tübingen

# Inhaltsverzeichnis

1. Einleitung.....	3
2. Bestimmende Haltungen der Sprachpolitik.....	3
3. Strömungen in Europa.....	4
4. Sprachpolitik Japans.....	4
5. Die Netz-Wirklichkeit.....	5
5.1. Ermittlung von Daten.....	5
5.1.1. Probleme.....	5
5.1.2. Lösung.....	6
5.2. Hauptsprachen des Internets.....	9
5.3. Entwicklung.....	9
5.3.1. Stand 1997.....	10
5.3.2. Stand Anfang 2001.....	11
5.3.3. Stand Juni 2001.....	11
5.3.4. Stand Januar 2002.....	11
5.3.5. Stand 2002.....	12
5.3.6. Momentaufnahme 2005.....	12
5.3.7. Inhalte.....	13
5.4. Analyse.....	13
5.5. Prognose.....	14
5.6. Unterschiedliche Sprachnutzung.....	15
5.7. Aktivitäten im Netz.....	17
5.7.1. Übersetzungsdienste und Wörterbücher im Internet.....	17
5.7.2. Das Phänomen Wikipedia.....	18
5.7.3. Spezialfall Programmiersprachen.....	20
6. Politische Vorgaben und Streitthemen.....	21
7. Fazit.....	22
8. Literaturverzeichnis.....	23
Anhang A.....	25
Anhang B.....	26

# 1. Einleitung

In dieser Arbeit soll die Sprachpolitik, Entwicklung und Verbreitung der Sprachen im Internet untersucht werden.

Zunächst werden die grundlegenden Haltungen in der Sprachpolitik herausgearbeitet und die verschiedenen Strömungen in Europa benannt. In einem gesonderten Punkt wird auf die Sprachpolitik Japans eingegangen.

Anschließend wird untersucht, weshalb die Sprachverteilung im Internet nicht der der wirklichen Welt entspricht. Dabei wird auf die spezifische Problematik der Datenerhebung im Internet eingegangen, die Entwicklung der Sprachverteilung im Internet und eine Prognose angestellt. Im unmittelbarem Anschluss wird versucht, die Wirkung und den Zweck unterschiedlicher Sprachnutzung zu erfassen und Erklärungen für widersprüchliche, zuvor erhobene Daten zu finden.

Im letzten Teil dieser Arbeit wird anhand des aktuellen Streits um die Digitalisierung von Schriftgut eine zuvor postulierte Wechselbeziehung von Sprachpolitik und Aktivitäten im Netz untermauert. Schließlich werden in einem Fazit Vermutungen angestellt, ob es eine vielsprachige oder einsprachige Zukunft für das Netz geben wird.

## 2. Bestimmende Haltungen der Sprachpolitik

Es lassen sich in der Sprachpolitik acht Haltungen feststellen (vgl. Treanor):

- **Sprache als Übel:**

„Die Sprachenvielfalt ist ein Übel und eine Konfliktursache. Man sollte sie durch eine universelle Sprache oder zumindest durch eine globale Hilfssprache überwinden.“

- **Gegen Standardisierung:**

„Die sprachliche Standardisierung in der Moderne ist, wie die Newspeak in Orwells '1984', unmenschlich und von Übel.“

- **Für die Vielfalt:**

„Sprachliche Vielfalt ist ein Wert an sich und mit der biologischen Vielfalt vergleichbar.“

- **Schutz der Sprache:**

„Jede Sprache hat einen Wert an sich und sollte, vielleicht wie ein Kunstwerk, geschützt werden.“

- **Bedeutung der Sprache:**

„Für Menschen und Nationen sind Sprachen eine wesentliche Grundlage. Die Sprache ist mit Identität, Kultur und Gedächtnis verbunden. Ein Sprachzerfall ist ein kultureller Genozid.“

- **Universalitätsanspruch: Religiöse Bedeutung**

„Eine bestimmte Sprache ist allen anderen überlegen, weil sie Wahrheit, Wert oder das Wort Gottes zum Ausdruck bringt. Sie sollte universell, vielleicht nur für religiöse Zwecke, verwendet werden.“

- **Universalitätsanspruch: Politik und Gesellschaft**

„Eine bestimmte Sprache ist mit einer überlegenen politischen Philosophie oder einer überlegenen Gesellschaftsform verbunden. Sie sollte zur universellen Sprache werden.“

- **Konformität:**

„Eine universelle oder nahezu universelle Sprache, wie es sie in der Vergangenheit gegeben hat, sollte wieder hergestellt werden.“

### 3. Strömungen in Europa

Obgleich Europa von Manchen als das „Sprachen-Eldorado“ (Hashimoto, 2005) bezeichnet wird, wirft die Existenz vieler unterschiedlicher Sprachen in Europa die Frage auf, wie Kommunikation stattfinden und verbessert werden kann. Es lassen sich verschiedene Strömungen in Europa ausmachen (vgl. Treanor):

<i>unterschiedliche Sprachen</i>	<i>einheitliche Sprache</i>
Nationalisten	Kosmopologen
Regionalisten	

Tabelle 1 Unterschiedliche Strömungen in Europa

Unterschiedliche Sprachen werden sowohl von den *Nationalisten*, die die Nationalsprachen unterstützen, als auch den *Regionalisten*, welche die regionalen Sprachen fördern, befürwortet. Allen beiden ist gemeinsam, dass sie, anders als die *Kosmopologen*<sup>1</sup>, für Vielsprachigkeit eintreten.

Es lässt sich eine weitere Unterscheidung durchführen:

- Neo-Atlantiker: Sie wollen Englisch als Kontaktsprache.
- Verteidiger der Nationalsprachen: Sie sind für eine begrenzte Vielsprachigkeit.
- technologische Optimisten: Sie sind der Ansicht, dass „bald“ völlig automatisierte Übersetzung zwischen den Sprachen technisch möglich wird.

Interessanterweise lässt sich hier eine weitere Gruppe erkennen, nämlich die der *Optimisten*. Sie unterscheiden sich dadurch von den anderen, dass sie implizit *jede* Sprache unterstützen. Tabelle 1 muss also erweitert werden:

<i>unterschiedliche Sprachen</i>		<i>einheitliche Sprache</i>
<b>begrenzt</b>	<b>unbegrenzt</b>	Kosmopologen
Nationalisten	Regionalisten	

Tabelle 2 Unterschiedliche Strömungen in Europa - erweiterte Darstellung

### 4. Sprachpolitik Japans

Wie steht es um die Sprachpolitik Japans? Das EURAC-Institut bezeichnet Japan als das „Land des Monolinguisimus“ (vgl. Hashimoto, 2005). Aus einem Interview mit Satoshi Hashimoto, Dozent für

<sup>1</sup> Von Treanor als „Befürworter einer kosmopolitischen Ethik“ bezeichnet

europäische Klein- und Minderheitensprachen an der Universität von Hokkaido, lässt sich Folgendes über die japanische Sprachpolitik entnehmen (vgl. Hashimoto, 2005):

- Regionalsprachen werden in Japan nicht gefördert
- „auf akademischer Ebene gibt es höchstens 20 Experten für europäische Sprachen“
- „ethnische Minderheiten in Japan werden nicht anerkannt“, ihre Sprache nicht unterrichtet
- „Minderheitensprachen werden ganz bewusst unterdrückt“
- „die einzig wirklich anerkannte und geförderte Fremdsprache ist Englisch“

Hashimoto bezeichnet sein Land in sprachpolitischer Hinsicht als „auf dem besten Weg in die monolinguische Sackgasse“. Auf diese Weise geht auch Verständnis für die eigene Geschichte verloren, beispielsweise bedeutet „Sapporo“, der Name der Hauptstadt von Japans zweitgrößter Insel, in der Sprache Ainu der Eingeborenen Japans, „Land, das der Fluss gab“. Dies ist jedoch nur den „wenigsten“ Japanern bekannt, und da diese Sprache vom Aussterben bedroht ist, wird es möglicherweise irgendwann niemand mehr wissen.

## **5. Die Netz-Wirklichkeit**

Bisher wurde die allgemeine Sprachpolitik Europas und Japans besprochen. Diese wirkt sich auf die Bereiche aus, auf die die Politik unmittelbaren Einfluss oder die Möglichkeit zur Kontrolle hat, wie etwa die Sprachnutzung in Politik und Wirtschaft oder der Sprachunterricht an Schulen. Das Internet jedoch ist als globales Kommunikationsmittel nur sehr beschränkt diesem Einfluss ausgesetzt. Spiegelt es dennoch den Zustand der nichtvirtuellen Welt wider? Oder unterscheidet es sich, stellt vielleicht gar einen Gegenpol dar? Im Folgenden soll nun die Wirklichkeit im Netz in seiner Entwicklung untersucht, analysiert und sofern möglich auch prognostiziert werden.

### **5.1. Ermittlung von Daten**

#### **5.1.1. Probleme**

Leider ergeben sich bei der Untersuchung der Netz-Wirklichkeit verschiedenartige Probleme. Zunächst muss geklärt werden, was denn die „Netz-Wirklichkeit“ sein soll. Während über die reale Welt auf publizierte Statistiken, die z. B. die Zahl der unterrichteten Schüler in einer bestimmten Fremdsprache, die Zahl der veröffentlichten Bücher in einer Sprache oder die Einwohnerzahl eines Landes mit einer bestimmten Sprache aufführen, zurückgegriffen werden kann, existieren für das Internet kaum verlässliche Statistiken. Das hat mehrere Ursachen:

- Jede Person kann jederzeit, abgesehen von technischem Problemen oder einer gesetzlichen Zugangsbeschränkung, unkontrolliert selbstständig neue Inhalte ins Netz einstellen oder vorhandene verändern.
- Es gibt keine zentrale Stelle, die eine genaue Übersicht über alle bereits existenten Inhalte im Netz hat. Und selbst wenn es sie gebe, wären ihre Informationen bereits nach wenigen Sekunden veraltet, da sich Inhalte im Netz fast in Echtzeit verändern lassen.

- Aus wirtschaftlichen, rechtlichen oder politischen Gründen können technische Zugangsbeschränkungen für bestimmte Personenkreise oder Institutionen eingesetzt werden. Ein Beispiel hierfür wäre die kostenlose Distribution eines Computerspieles für Werbezwecke, die jedoch nur für die eigene Bevölkerung gelten soll.

Insbesondere das mangelnde Wissen, welche Inhalte wo im Netz zu finden sind, macht das Erstellen einer aussagekräftigen Statistik schwierig. Eine Befragung sämtlicher Menschen und Unternehmen mit direktem oder auch nur indirektem Internetzugang scheidet aus praktischen Gründen aus.

### 5.1.2. Lösung

Es bleibt also nur die Möglichkeit, selbstständig nach den Inhalten zu suchen. An dieser Stelle kommen die *Suchmaschinen* ins Spiel. Sie werden fast ausschließlich von Unternehmen betrieben und durchforsten das Internet nach möglichen Inhalten. Dabei sind sie, je nach eingesetzter Technik, unterschiedlich erfolgreich: So gab zum Beispiel die Suchmaschine „Google“ an, 8.058.044.651 Seiten<sup>2</sup> (ca. acht Milliarden Seiten) (vgl. Google, 2005) indiziert zu haben. Konkurrenten geben jedoch wie im Fall von „Gigablast“ mit 2.060.245.584 Seiten (ca. zwei Milliarden Seiten) (vgl. Gigablast, 2005) oder „Yahoo“ mit ca. 19 Milliarden Seiten (vgl. Sullivan, 2005) ganz andere Zahlen wieder, die Ergebnisse sind also nicht annähernd vergleichbar. Allein schon darüber, was denn überhaupt eine „Seite“ ist, lässt sich streiten. Ist zum Beispiel eine Seite, die nur technische Befehle oder Beschreibungen für den Internetbrowser und keine für den Betrachter sicht- oder lesbare Information enthält, eine Seite, die einen Inhalt enthält? Wenn nein, was ist mit einer Seite, die nur ein Bild enthält, möglicherweise eine der weit verbreiteten Baustellengrafiken, die lediglich angeben, dass sich „noch nichts“ auf dieser Seite befindet? Was ist mit einem künstlerischen Internetauftritt, bei dem sich der Benutzer durch mehrere Seiten bewegt, um schließlich auf eine bewusst leer gelassene Seite zu treffen, also quasi „ins Leere fällt“? An und für sich stellt dies ebenfalls einen Inhalt, nämlich einen *Nichtinhalt*, dar. Muss die Suchmaschine dies analysieren? *Kann* sie es überhaupt?

Zunächst soll akzeptiert werden, dass es unterschiedliche Auffassungen über die Definition einer Seite gibt. Schließlich ist dies nicht der zentrale Gegenstand der Untersuchung, wie die Wirklichkeit in Bezug auf Sprachpolitik im Netz aussieht. Da Inhalte wie oben beschrieben noch nicht zuverlässig automatisch analysiert werden können (vgl. Korpela, 2003), wird im Folgenden für die Untersuchung der Sprachverteilung im Internet die Anzahl der Seiten mit maschinell auswertbaren, für den Betrachter sicht- und lesbaren Inhalten zugrunde gelegt.

Dies bringt jedoch das nächste Problem mit sich: Wie wird denn ermittelt, in welcher Sprache Inhalte auf einer Seite verfasst sind? Hierbei gibt es verschiedene Ansätze: Die Seite selbst zu analysieren oder möglicherweise in die Seite eingebaute spezielle Informationsanweisungen, so genannte *Meta-Tags*<sup>3</sup>, zu verwenden.

2 Leider gibt Google inzwischen die Anzahl der indizierten Seiten nicht mehr an, weswegen ein Vergleich mit aktuellen Zahlen nicht möglich ist (vgl. Sullivan, 2005). Allerdings dürfte sich bei der Größenordnung von Milliarden Seiten keine wirklich nennenswerte Verschiebung ergeben haben.

3 Ein *Tag* ist ein in Spitzklammern eingeschlossener Befehl, z.B. leitet das Tag „<p>“ einen neuen Absatz ein, der dann mit „</p>“ abgeschlossen wird. *Meta-Tags* sind eine besondere Gruppe von Tags: Sie haben keine Auswirkung auf die Gliederung oder Darstellung der Seite, sondern enthalten Zusatzinformationen.

<i>eigene Analyse</i>	<i>Seiteninformation</i>
Ländersuffix	<i>Meta-Tags</i>
Wörterbuchanalyse	
Koinzidenzindex	

Tabelle 3 Die unterschiedlichen Ansätze zur Sprachanalyse einer Internetseite mit ein paar Beispielverfahren

Meta-Tags sind, sofern sie vorhanden sind und richtig eingesetzt werden, die zuverlässigste Informationsquelle. Da sie jedoch beim Aufruf einer Seite mit übertragen werden müssen und somit zusätzliche Zeit für das Laden der Seite in Anspruch nehmen, werden sie unter Umständen „wegoptimiert“. Dies geschieht heutzutage aufgrund der zunehmenden Anzahl von Breitbandverbindungen zwar seltener, doch sind noch viele Seiten älteren Datums vorhanden, die für diese Auswertung nicht einfach übergangen werden dürfen. Auch müssen Meta-Tags vom Benutzer eingebaut werden, was einen zusätzlichen Aufwand bedeutet. In der Praxis kommt noch ein weiteres Problem hinzu, nämlich die Verwendung von Programmen zur Publizierung der Inhalte im Internet. Werden Programme zur Verwaltung und Generierung von Internetseiten wie „Microsoft Frontpage“ oder „Netobjects Fusion“ nicht in einer lokalisierten Version, sondern in der englischen Originalversion verwendet, werden diese davon ausgehen, dass die zu publizierenden Inhalte englische Inhalte sind und entsprechende Meta-Tags setzen. Wird die Seite dann ausschließlich anhand der Meta-Tags sprachlich eingeordnet, würde sie fälschlicher Weise als englische Seite eingestuft werden.

Wird jedoch auf die korrekte Setzung der Meta-Tags geachtet, so treten in der Praxis leicht Problemfälle auf, z. B. wie eine Seite ausgezeichnet werden soll, auf der sich Inhalte in mehreren Sprachen gleichzeitig befinden. Wie ist beispielsweise eine Seite zu behandeln, auf der ein englisches Gedicht im Original wiedergegeben, aber anschließend auf Deutsch diskutiert wird: Ist es nun eine deutsche Seite oder eine englische Seite? Technisch gibt es auch dafür eine Lösung: Die Seite kann mithilfe von *Tags* (s. o. *Meta-Tags*) in Bereiche unterteilt und mit *Attributen* genau einer Sprache zugeordnet werden. Dies ist jedoch ein noch viel größerer Aufwand als bereits die bloße Auszeichnung der Seite als Ganzes.

Verlässt man sich nicht auf die gelieferten Informationen, so muss die Seite bzw. ihr Inhalt analysiert werden. Dafür gibt es verschiedene Verfahren. Im Folgenden sollen nun drei sehr einfache vorgestellt werden: Ländersuffix, Koinzidenzindex und Wörterbuch<sup>4</sup>.

### Ländersuffix

Es wird einfach anhand der Adresse der Seite geprüft, welchem Land sie zuzuschreiben ist. Z. B. steht „de“ für Deutschland, „uk“ für Großbritannien und „jp“ für Japan. Anhand des Länderkürzels kann dann die Sprache zugewiesen werden.

In der Praxis ist dies jedoch kein sinnvolles Verfahren, da es Länder mit mehreren offiziellen oder gebräuchlichen Sprachen gibt (z. B. die Schweiz), und eine Seite nicht unbedingt Inhalte in einer der Adresse zuordbaren Sprache enthalten muss (z. B. ein mehrsprachiger Internetauftritt). Für sich allein

<sup>4</sup> Leider ließ sich nicht ermitteln, wie genau Suchmaschinen die Sprache einer Seite erfassen, da sie sich – wohl aus Gründen des Wettbewerbs – sehr verschlossen geben. Die angegebene Verfahren sind darum als wahrscheinlich und nicht verifiziert zu betrachten.

genommen ist es also unbrauchbar, doch als Ergänzung zu einem anderen Verfahren kann es durchaus nützlich sein, weswegen es hier auch aufgeführt ist.

### Koinzidenzindex

Mit dem aus der Kryptoanalyse stammenden *Koinzidenzindex* können Texte auf sprachliche Eigenschaften untersucht werden (vgl. Hauck, 2005: S. 31-36). Dabei wird berechnet, wie oft ein Buchstabe mit einem anderen Buchstaben gemeinsam vorkommt. Aus dieser Information lässt sich dann ein Wert berechnen, der für jede Sprache typisch ist, beispielsweise ist er ca. 0,0762 für Deutsch, ca. 0,0661 für Englisch und ca. 0,0385 für einen Text mit zufälligen Buchstabenkombinationen (vgl. Koinzidenzindex).

Um die Sprache einer Seite zu ermitteln, kann der Koinzidenzindex für diese Seite berechnet und anschließend mit den bekannten Werten für die verschiedenen Sprachen verglichen werden. Allerdings darf der Text der Seite dazu nicht zu kurz sein, da dies ebenso wie die Mischung mit einer anderen Sprache eine Verzerrung der Statistik bewirken würde.

Zu den Vorteilen des Koinzidenzindexes gehört, dass er sich sehr einfach berechnen lässt und keine großen Anforderungen an die verwendeten Systeme stellt. Zudem ist er für längere, wirklich einsprachige Texte sehr zuverlässig. Tatsächlich kann man mit ihm sogar die Sprache eines – mit einfachen Methoden – verschlüsselten Textes feststellen, selbst wenn man diesen nicht entschlüsseln kann (vgl. Hauck, 2005: S. 38).

### Wörterbuch

Bei der Verwendung eines Wörterbuchs lässt sich die Sprache einer Seite zuverlässig feststellen. Dazu muss eine Liste der Wörter für alle Sprachen existieren. Die Wörter im Text der Seite müssen dann lediglich nachgeschlagen und die Anzahl der Treffer je Sprache notiert werden. Schließlich wird der Seite die Sprache zugewiesen, die die meisten Treffer aufweist.

So einfach und nachvollziehbar dieses Verfahren ist, so hat es leider auch ein paar Nachteile. Einerseits wären da die großen Anforderungen an die Systeme, die die Sprache der Seite ermitteln sollen: Wörterbücher für jede Sprache müssen vorgehalten werden und jedes Wort in jedem der Wörterbücher nachgeschlagen werden, was bei Milliarden von Seiten (s. o.) und Hunderten von Sprachen<sup>5</sup> einen zwar automatisierbaren, aber dennoch immensen Aufwand bedeutet. Leider kann nicht abgebrochen werden, nachdem das nachzuschlagende Wort in einer Sprache gefunden wurde, da die meisten Sprachen Fremdwörter enthalten, z.B. sind die eigentlich englischen Begriffe „Download“ oder „Link“ auf zahlreichen deutschen Seite zu finden. Noch viel problematischer ist es natürlich auch wieder, wenn die Seite selbst wie beim Beispiel des englischen Gedichts mit deutscher Diskussion oben verschiedene Sprachen enthält.

Ein weiteres Problem, das bei der Verwendung von Wörterbüchern auftritt, sind Rechtschreibfehler auf der Seite oder die Verwendung von Begriffen, die nicht im Wörterbuch enthalten sind, beispielsweise „downgeloadet“, „gedownloadet“, „Babies“, „Parties“, regional gebräuchliche Begriffe wie „Semmel“ usw.

---

<sup>5</sup> Der „Ethnologue language name index“ gibt für die Anzahl der primären Sprachennamen 7.299, die Anzahl der alternativen und dialektischen Sprachen sogar 39.491 an (vgl. Gordon, 2005)

## Kombination

Keines der oben aufgeführten Verfahren ist für sich selbst genommen ausreichend. Denkbar ist daher eine Kombination, beispielsweise eine Schnellbestimmung der möglichen Sprachkandidaten mittels des Ländersuffixes und des Koinzidenzindex mit anschließender Wörterbuchanalyse. Sofern sich keine der Sprachen als deutlich wahrscheinlicher erweist, können die Meta-Tags zur ausschlaggebenden Bewertung herangezogen werden.

Allerdings kann selbst bei der Kombination verschiedener Verfahren nicht von einer sicheren Erkennung gesprochen werden. Lediglich eine Wahrscheinlichkeitsaussage kann getroffen werden, weshalb alle im Folgenden aufgeführten Daten eine unbekannte Fehlerrate enthalten. Stichproben haben jedoch ergeben, dass die Fehlerrate<sup>6</sup> bei der Erkennung von Deutsch 7,8%, Englisch 5,5% und Französisch 3,0% betragen könnte (vgl. Netz-Tipp, 2000).

Sich diese Problematik zu verdeutlichen ist von elementarer Bedeutung. Anhand der zuvor aufgeführten Verfahren sollte auch noch einmal darauf hingewiesen werden, auf wievielen ungesicherten Informationen die weiteren Untersuchungen und Überlegungen in dieser Arbeit basieren.

## **5.2. Hauptsprachen des Internets**

Im Folgenden wird nun die Entwicklung der Sprachverteilung im Internet dargestellt<sup>7</sup>. Sie lässt sich kurz zusammenfassen (vgl. Netz-Tipp, 2000<sup>8</sup>):

- klar dominierend: Englisch
- überraschend: Die „Zweitsprache“ im Netz ist Deutsch
- abnehmend: Japanisch

## **5.3. Entwicklung**

Für die Jahre 1997 bis 2002 lassen sich im Internet publizierte Statistiken finden (vgl. Netz-Tipp, 2000):

---

6 Eine Fehlerrate von im Falle von Deutsch 7,8% bedeutet, dass 7,8% aller scheinbar deutschen Texte anderssprachig zu sein scheinen (vgl. Netz-Tipp, 2000).

7 Leider ist nur in einem einzelnen Fall die Angabe vorhanden, ob sich die Daten lediglich auf HTML-Daten oder auch auf andere Formate wie PDFs oder Microsoft Word DOC Dateien beziehen. Hier wird angenommen, dass die Angaben für HTML-Daten gelten.

8 Zwar steht auf der Seite „© 2000“, jedoch sind einige der präsentierten Zahlen neueren Datums, weswegen dieses Veröffentlichungsdatum falsch sein muss.

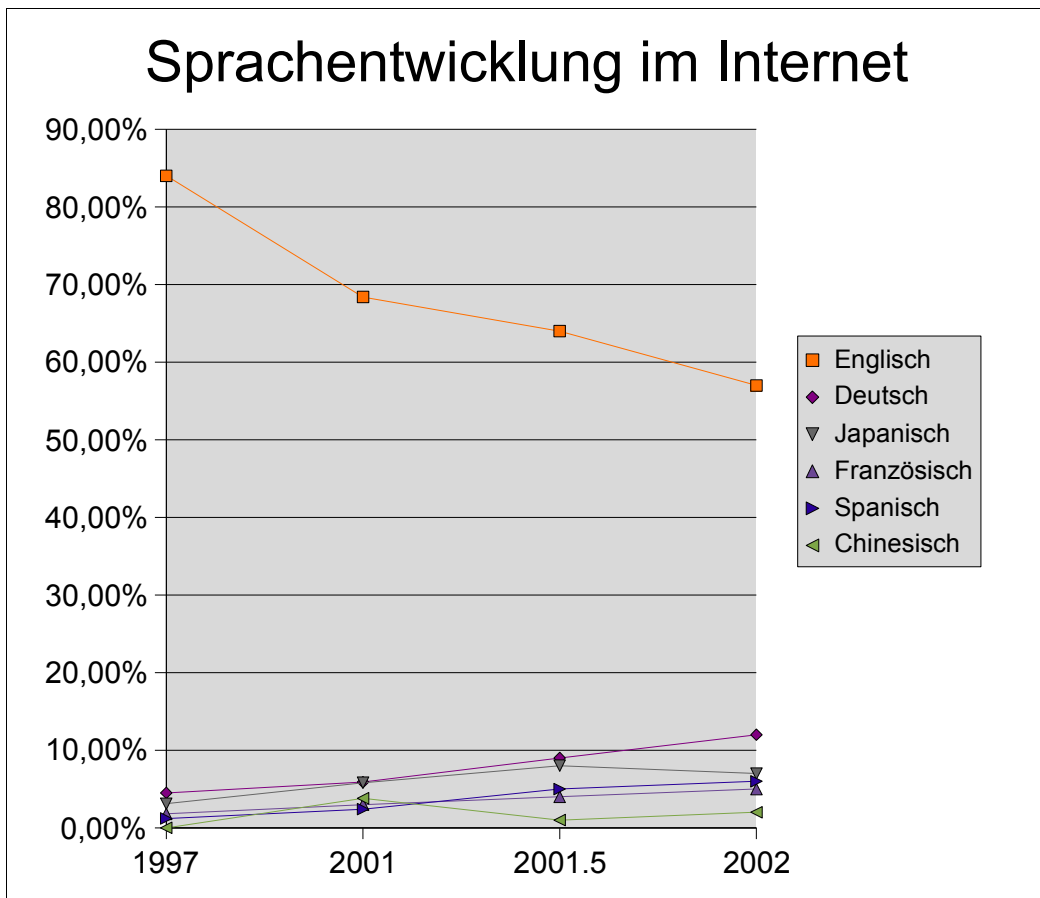


Abbildung 1 Entwicklung der Sprachverteilung im Netz. Quelle: Netz-Tipp, 2000

Diese Entwicklungen im Detail:

#### 5.3.1. Stand 1997

<i>Sprache</i>	<i>Anteil</i>
Englisch	84,00%
Deutsch	4,50%
Japanisch	3,10%
Französisch	1,80%
Spanisch	1,20%
Chinesisch	---

Tabelle 4 Sprachverteilung 1997. Quelle: Web Languages Hit Parade, 1997

### 5.3.2. Stand Anfang 2001

<i>Sprache</i>	<i>Anteil</i>
Englisch	68,40%
Japanisch	5,90%
Deutsch	5,80%
Chinesisch	3,80%
Französisch	3,00%
Spanisch	2,40%

Tabella 5 Sprachverteilung Anfang 2001. Quelle: Netz-Tipp, 2000

### 5.3.3. Stand Juni 2001

<i>Sprache</i>	<i>Anteil</i>
Englisch	64,00%
Deutsch	9,00%
Japanisch	8,00%
Spanisch	5,00%
Französisch	4,00%
Italienisch	2,00%
Chinesisch	1,00%
(Andere)	4,00%

Tabella 6 Sprachverteilung im Juni 2001. Quelle: Netz-Tipp, 2000

### 5.3.4. Stand Januar 2002

<i>Sprache</i>	<i>Anteil</i>
Englisch	57,00%
Deutsch	12,00%
Japanisch	7,00%
Spanisch	6,00%
Französisch	5,00%
Italienisch	3,00%
Chinesisch	2,00%
(Andere)	8,00%

Tabella 7 Sprachverteilung im Januar 2002. Quelle: Netz-Tipp, 2000

### 5.3.5. Stand 2002

<i>Sprache</i>	<i>Anteil</i>
Englisch	56,40%
Deutsch	7,70%
Französisch	5,60%
Japanisch	4,90%
Spanisch	3,00%
Chinesisch	2,40%

Tabelle 8 Sprachverteilung 2002. Quelle: Netz-Tipp, 2000

### 5.3.6. Momentaufnahme 2005

Eine Anfrage nach Seiten im Internet am 28.05.2005 um 21:20 Uhr bei Google ergab bei einer Gesamtzahl von 8,058,044,651 Seiten, dass Englisch einen Anteil von ca. 42% (3,380,000,000 Seiten) hat während Deutsch auf nur ca. 1% (76,300,000 Seiten) kommt. Gesucht wurde mittels Setzen des Sprachfilters von Google auf „Englisch“ bzw. „Deutsch“ und der Verwendung von einzelnen Buchstaben des Alphabets wie „e“ als Suchbegriff.

Die Abwärtsentwicklung von Englisch scheint sich zu bestätigen. Verwirrend ist jedoch der dem sonstigen Trend widersprechende geringe Anteil von Deutsch, der entweder auf ein technisches Problem oder auf das oben besprochene Sprachenidentifikationsproblem hinweist.

Tatsächlich ergeben sich bei genauerer Untersuchung sehr unterschiedliche Ergebnisse je nach verwendetem Suchausdruck. So führte erneutes Suchen mit Google am 27.10.2005 zwischen 13:20 Uhr und 13:30 Uhr bei der Suche nach deutschen Seiten über *oder*-verknüpfte Suchanfragen zu folgenden Ergebnissen:

<i>Suchbegriffe</i>	<i>Treffer</i>	<i>Anteil in %<sup>9</sup></i>
und	223.000.000	2,77%
sind	116.000.000	1,44%
auch	120.000.000	1,49%
zu	171.000.000	2,12%
und ODER sind	189.000.000	2,35%
sind ODER auch	189.000.000	2,35%
und ODER sind ODER auch	189.000.000	2,35%
auch ODER zu	189.000.000	2,35%

Tabelle 9 Unterschiedliche Ergebnisse für die Suche nach häufigen deutschen Wörtern mittels Google

Die Unterschiede zwischen den einzelnen Ergebnissen sind frappierend. Am schlüssigsten scheint noch 2,77% zu sein, doch selbst diese Zahl ist angesichts der bisherigen Entwicklung sehr unwahrscheinlich.

<sup>9</sup> Wegen mangelnder aktueller Daten (vgl. Sullivan, 2005) wird als Gesamtzahl die letzte sichere Angabe vom 28.05.2005 mit 8,058,044,651 Seiten zur Berechnung des prozentualen Anteils verwendet.

Bestätigt wird dies durch eine Suche nach Seiten in Englisch, die das deutsche Wort „und“ enthalten: Google liefert 73.000.000 Treffer, von denen augenscheinlich alles deutsche Seiten sind. Diese Fehlerrate ist enorm, hebt sie doch den deutschen Anteil im Netz um mindestens einen weiteren Prozentpunkt an.

### 5.3.7. Inhalte

Eine interessante Auskunft ist für das Jahr 2002 gegeben (vgl. Netz-Tipp, 2000): Die Sprachverteilung für PDF-Dokumente. Dies ist von besonderer Bedeutung, da PDF-Dateien anders als HTML-Seiten in der Regel einen hohen Informationsgehalt haben, also stärker als HTML-Seiten zu gewichten sind. Hierbei zeigen sich deutliche Unterschiede zu den anderen Statistiken, so hat hier Englisch einen deutlich niedrigeren Anteil und Japanisch einen um +2,5%, Deutsch einen um +3%, und Französisch einen um +3,2% höheren Anteil:

<i>Sprache</i>	<i>Anteil</i>
Englisch	38,17%
Deutsch	10,64%
Französisch	8,80%
Japanisch	7,41%
Spanisch	5,81%
Chinesisch	2,80%

Tabelle 10 Sprachverteilung von PDF-Dateien im Netz. Stand 2002. Quelle: Netz-Tipp, 2000

## 5.4. Analyse

Werden die oben aufgeführten Daten als Grundlage genommen, so lässt sich feststellen, dass die Verteilung der Sprachen im Internet nicht der weltweiten Verteilung der realen Welt entspricht (vgl. Global Reach [1], 2004). Schwierig wird es aber, wenn die Ursachen dafür ergründet werden sollen.

So wird vermutet, dass z. B. die starke Stellung des Deutschen mit einer besseren wirtschaftlichen Situation und Bildung derer zusammenhängt, die Deutsch als Fremdsprache sprechen (vgl. Netz-Tipp, 2000). Auch ist Deutsch wie Englisch, Französisch und Spanisch eine Verkehrssprache, wird also auch von Menschen mit anderer Muttersprache genutzt.

Anders verhält es sich mit dem Sinken des japanischen Anteils. Der postulierte Zusammenhang mit dem Schwinden des technologischen Vorsprungs erscheint eher wage, wie folgende Nutzerstatistik zeigt (vgl. Global Reach [2], 2004):

<i>Sprache</i>	<i>derzeitige Nutzer (in Mio)</i>	<i>potenzielle Nutzer (in Mio)</i>	<i>Ausschöpfungsgrad</i>
Japanisch	105	125	84,00%
Deutsch	71	98	72,00%
Französisch	49	72	68,00%
Englisch	300	508	59,00%
Chinesisch	220	885	25,00%

Tabelle 11 Ausschöpfungsgrad potenzieller Internetnutzer verschiedener ausgewählter Sprachen. Stand 30.03.2004. Quelle: Global Reach [2], 2004

Auch wenn Deutschland aufgeholt hat, so sind doch immerhin noch 34 Millionen Japaner mehr online als Deutsche. Dies sollte sich in einem deutlich größeren Anteil an japanischen Inhalten im Netz äußern. Geschwundener technologischer Fortschritt kann also höchstens eine von mehreren Ursachen sein.

## 5.5. Prognose

Die Entwicklung der Sprachen im Internet scheint in Richtung Vielsprachigkeit zu weisen (vgl. Netz-Tipp, 2000). Als wichtiger Faktor wird die *usability*, die Brauchbarkeit einer Internetseite ausgemacht. Dies betrifft, abgesehen von optischen Kriterien wie ein klar gegliederter Seitenaufbau oder passend gewählte Farbkombinationen, insbesondere die Sprache des Inhalts der Seite. Insbesondere für Unternehmen, die ihre Dienste im Internet anbieten, kann es sich auszahlen, in die Vielsprachigkeit ihres Internetauftritts zu investieren. So bietet die amerikanische Suchmaschine Google auf ihrer deutschsprachigen Seite „zuweilen klareres Deutsch (und ist auch im deutschen Sprachraum erfolgreicher) als die deutsche Konkurrenz“ (Netz-Tipp, 2000).

Es wird vermutet, dass vor allem die Sprachen Englisch, Deutsch, Französisch, Japanisch, Spanisch und Chinesisch im Netz weit verbreitet sein werden (vgl. Netz-Tipp, 2000). Gestützt wird diese These durch das Unternehmen Global-Reach, welches insbesondere die Zunahme von anderen Sprachen als Englisch dokumentiert und prognostiziert (vgl. Global Reach [2], 2004):

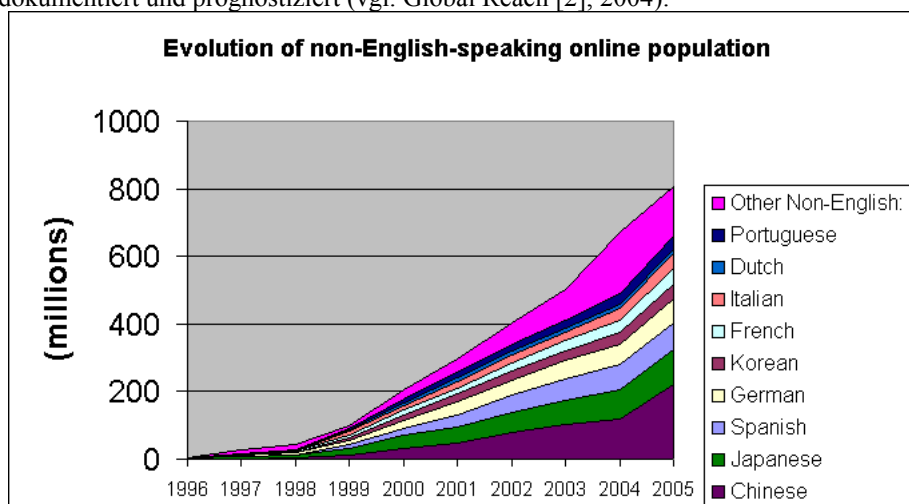


Abbildung 2 Entwicklung der Internetnutzer, die nicht Englisch sprechen. Stand: 30.03.2004. Quelle: Global Reach [2], 2004

Regionale Sprachen wird es für spezielle Angebote geben, das Aussterben „kleiner“ Sprachen wird vom Internet nicht beeinflusst werden (vgl. Netz-Tipp, 2000). So ließen sich beispielsweise ca. 71.900 friesische Seiten und ca. 30.000 lateinische Seiten<sup>10</sup> finden.

Allerdings lässt sich auch die These aufstellen, dass das Internet das Aussterben kleiner Sprachen aufhalten kann. Die zuvor genannten friesischen und lateinischen Seiten belegen, dass Inhalte auch in diesen Sprachen weiterhin publiziert werden, und sich die Nutzer von derartigen Angeboten im Internet zu neuen Gemeinschaften zusammenfinden können. Ein weiteres Beispiel hierfür ist die Seite „Ohwejagehka: Ha` degaenage“, die den Erhalt der Ogwehoweh-Sprachen zum Ziel hat (vgl. Ogwehoweh, 2003).

<sup>10</sup> Eine hohe Fehlerquote erforderte viele Korrekturen (vgl. Netz-Tipp, 2000).

Auffällig ist auch, um wie viel schwächer die Stellung des Englischen ist, als zunächst vermutet werden könnte. So werden der Möglichkeit, dass Englisch sich als *lingua franca* des Netzes dauerhaft etablieren könnte, klare Absagen erteilt (vgl. Horvath, 1997; vgl. Korpela, 2003): Englisch wäre zu schwierig, zu umfangreich und im Zerfall begriffen – „ein natürlicher Prozess“ (Horvath, 1997), den auch schon andere Sprachen wie Latein durchlaufen haben.

## 5.6. Unterschiedliche Sprachnutzung

Werden auf einer Seite die Inhalte in mehreren Sprachen angeboten, so heißt das nicht unbedingt, dass die Inhalte die gleichen sind. Dies lässt sich besonders gut anhand des folgenden Beispiels illustrieren: Zwei Bildschirmschnappschüsse des Änderungsprotokolls des japanischen Internetbrowsers „KazeHakase“ werden gezeigt, zuerst in der englischen Version.

The screenshot shows the website for KazeHakase, a Japanese internet browser. The page is titled "# Kazehakase-0.1.9 release" and is in English. The navigation menu includes: Home, About, Manual, Download, Mailing List, Bug Tracking, CVS, Documents (Wiki), and Antenna. The main content area is divided into three sections: "Recent Update RSS" with a list of updates from 0.1.2 to 0.2.3; "Changes" for version 0.1.9, listing features like thumbnail support, recent closed tabs, and proxy toggle; and "Notice" regarding the experimental state of thumbnail support. The page footer indicates the last update was on August 15, 2005 at 13:25.

Abbildung 3 Das Änderungsprotokoll eines japanischen Internetbrowsers namens KazeHakase. Hier die englische Version.  
Quelle: Kazehakase, 2005

Nun folgt die japanische Originalversion. Der Unterschied, auf den die Aufmerksamkeit hier gelenkt werden soll, ist (rot) markiert.

風博士 [English]

風博士について|ダウンロード|メーリングリスト|バグ|要望|CVS|Wiki| トップ «00018 最新 00020»  
アンテナ

Recent Update RSS

- 風博士 0.3.1 リリース
- 風博士 0.2.9リリース
- 風博士 0.2.7 リリース
- 風博士 0.2.6 リリース
- 風博士 0.2.5リリース
- 風博士 0.2.4リリース
- 風博士 0.2.3リリース
- 風博士 0.2.2リリース
- 風博士 0.2.1リリース
- 風博士 0.2.0リリース

Archives:  
00031 Go

Latest Screenshot

Aug. 28, 2004  
IPA様より, 2004 年度 未踏ソフトウェア創造事業として, 開発資金の援助を受けております。

# 風博士 0.1.9 リリース

今月の肉知識  
好きな動物は牛や豚。

変更点

- サムネイルサポート (Hiroyuki Ikezoe)
- 最近閉じたタブ (Hiroyuki Ikezoe)
- プロキシ使用/未使用 アクシオン (Hiroyuki Ikezoe)
- 履歴内全文検索用オプション追加 (Hiroyuki Ikezoe)
- Undo/Redo (Hiroyuki Ikezoe)
- Mozilla 1.7.2と1.8a3サポート (Hiroyuki Ikezoe)
- 詳細設定画面でページ内検索をすると落ちるバグの修正 (Hidetaka Iwai)

注意

サムネイルはまだ実験段階のため、デフォルトでは作成されません。サムネイルを作成するには、~/kazehakase/kzrc内の[Global]セクションに"create\_thumbnail=true"と書いてください。さらに、サムネイルを作成するためには、psファイルをpngに変換できるconvert(ImageMagickに付属)が必要です。

Last update on September 19, 2004 08:50

コメント(2) [コメントを投稿する]

- \_ doctor [gentooに関する報告はここでもよろしいのでしょうか...? kazehakase-0.1.9.ebuild を eme.]
- \_ zoe [etc-updateしてください。]

更新  
SOURCE RGE JP

Generated by tDiary version 2.1.3  
Powered by Ruby version 1.8.2

Abbildung 4 Das Änderungsprotokoll eines japanischen Internetbrowsers namens KazeHakase. Hier die japanische Originalversion. Quelle: Kazehakase, 2004

Die Informationen außerhalb des roten Bereichs sind weitgehend identisch, doch der rot markierte Teil fehlt in der englischen Fassung. Dort steht „今月の肉知識 ... 好きな動物は牛や豚”

(„Kongetsu no niku chishiki ... sukina dōbutsu wa ushi ya buta“), was soviel heißt wie „Die Fleischinformation dieses Monats: Ich mag Rind oder Schwein“.

Für den Nutzer des Internetbrowsers ist diese Information sicherlich unwichtig, doch zeigt es, dass Sprache unterschiedlich verwendet werden kann. Als Grund für dieses Verhalten ist denkbar:

- **Abgrenzung der Nutzergruppe:**

Es dient dem Ausdruck oder dem Wunsch nach einer persönlicheren Beziehung zum japanischen Nutzer, da der Autor etwas von sich persönlich erzählt. Dem englischen Nutzer fühlt er sich offenbar weniger verbunden.

- **Probleme bei der Übersetzung:**

Dem Auslassen könnte eine Unsicherheit im Umgang mit der englischen Sprache zugrunde liegen.

- **Fremdsprache ist formal:**

Eine Fremdsprache wird als formaler empfunden, weshalb derlei private Information dort nichts zu suchen hat.

Das bewusste Übersetzen oder Nichtübersetzen von Inhalten kann also verschiedenen Zwecken dienen. Tatsächlich lässt sich so im Allgemeinen steuern, wer welche Seite besucht, da ein englischer Internetnutzer sich wahrscheinlich eher die englische Fassung anstatt die japanische ansehen wird. Sprache kann also nicht nur dem Bereitstellen von Informationen dienen, sondern auch der Steuerung des Informationsflusses.

## **5.7. Aktivitäten im Netz**

Wie schon bei der Analyse angesprochen lassen sich nicht alle Unterschiede der Netz-Wirklichkeit zur realen Welt durch politische Maßnahmen oder ökonomische, technologische und soziale Faktoren erklären. Es müssen also weitere Gründe vorhanden sein.

Als ein möglicher Grund soll hier die Aktivität der Nutzer im Netz angeführt werden, die eine bestimmte Sprache sprechen. Illustriert und belegt werden soll dies an den Beispielen freier Wörterbücher im Internet, der Wikipedia und dem Spezialfall Programmiersprachen.

### **5.7.1. Übersetzungsdienste und Wörterbücher im Internet**

Frei verfügbare Wörterbücher und kostenlose Übersetzungsdienste scheinen zunächst von nicht besonders großer Bedeutung zu sein, da schließlich für so gut wie jede Sprache Wörterbücher erwerbbar sind. Zudem kann über die Qualität von Onlinewörterbüchern, wenn sie von Freiwilligen erstellt sind, kein generelles Urteil gefällt werden.

Beispielsweise ist das japanisch-deutsche Onlinewörterbuch „Wadoku“ (<http://www.wadoku.de>) mit 206.280 Einträgen (27.10.2005) außerordentlich umfangreich. Doch schwankt die Qualität der Einträge zwischen „sehr gut“ und „falsch“<sup>11</sup>. Allerdings bedeutet die freie, elektronische Verfügbarkeit, dass die Hürde, „schnell mal eben etwas nachzuschlagen“ sehr niedrig ist. Es gibt Erweiterungsprogramme wie „Moji“ für Internetbrowser oder eigenständige Programme wie „KITEN“ oder „JGLOSS“, die im Laufe ihrer mehrjährigen Entwicklung immer komfortabler geworden sind.

Mit sehr geringem Suchaufwand lassen sich Wörterbücher für Englisch, Deutsch, Spanisch, Französisch und Japanisch finden. Sucht man jedoch nach einer Sprache wie z.B. Estnisch erhält man nur

---

<sup>11</sup> Dieser Aussage basiert auf Stichproben mit verschiedenen gewöhnlichen und ungewöhnlichen Einträgen.

eine Liste kommerzieller Übersetzungsdienste. Dies hat zur Folge, dass Inhalte in dieser Sprache nicht im Original von Nichtmuttersprachlern gelesen werden können und die Kommunikation erschwert wird. Es ist also somit an dem Inhaltsschaffendem, eine Übersetzung anzufertigen und zu veröffentlichen. Im Falle eines Unternehmens ist es beispielsweise sehr wahrscheinlich, dass ausländische Interessenten sich an Konkurrenzunternehmen wenden, wenn deren Seiten übersetzt sind und die des eigenen Unternehmens nicht, da der Aufwand niedriger ist. Dies stellt somit eine wirtschaftliche Belastung dar.

Wie nützlich derart verfügbare Onlinewörterbücher sind ist von der Europäischen Union schon lange erkannt. So existiert seit 1973<sup>12</sup> das „Eurodicautum“ (<http://europa.eu.int/eurodicautom/Controller>), ein kostenlos nutzbares Fachbegriffswörterbuch. Es umfasst 5,5 Millionen Einträge in den Sprachen Dänisch, Finnisch, Griechisch, Portugiesisch, Holländisch, Französisch, Italienisch, Spanisch, Englisch, Deutsch, Latein und Schwedisch und wird 120.000 mal pro Tag verwendet. Allerdings fehlen hier acht der Sprachen, in denen sich das zentrale Onlineportal der Europäischen Union (<http://www.europa.eu.int/>) präsentiert, es besteht also offensichtlich noch Nachholbedarf.

### 5.7.2. Das Phänomen Wikipedia

Ein besonderes Augenmerk verdient das Projekt „Wikipedia“ (<http://www.wikipedia.org>). Wikipedia ist eine freie Enzyklopädie, die in Gemeinschaftsarbeit von 13.000 Freiwilligen geschaffen wurde und fortwährend erweitert und verbessert wird (vgl. Wikipedia [1], 2005). Von zentraler Bedeutung für die Wikipedia ist, dass sämtliche Artikel unter Angabe der Quelle und der Autoren frei verwendet und modifiziert werden dürfen. Die englische Wikipedia ging am 15.01.2001 online, der deutsche Ableger folgte bereits im März 2001 (vgl. Wikipedia [2], 2005) und der japanische im Mai 2001 (vgl. Wikipedia [1], 2005).

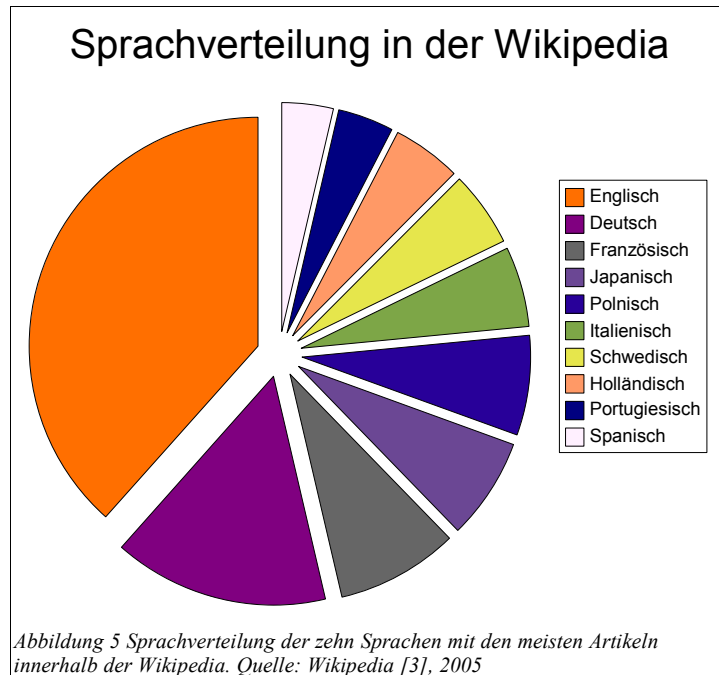
Für die Zwecke dieser Arbeit ist interessant, dass Wikipedia sehr viele Statistiken veröffentlicht. Zunächst einmal einige Informationen über die Sprachverteilung der 10 Sprachen mit der größten Anzahl von Artikeln:

<i>Sprache</i>	<i>Anzahl Artikel</i>
Englisch	778.000
Deutsch	305.000
Französisch	177.000
Japanisch	148.000
Polnisch	139.000
Italienisch	114.000
Schwedisch	110.000
Holländisch	101.000
Portugiesisch	80.000
Spanisch	71.000

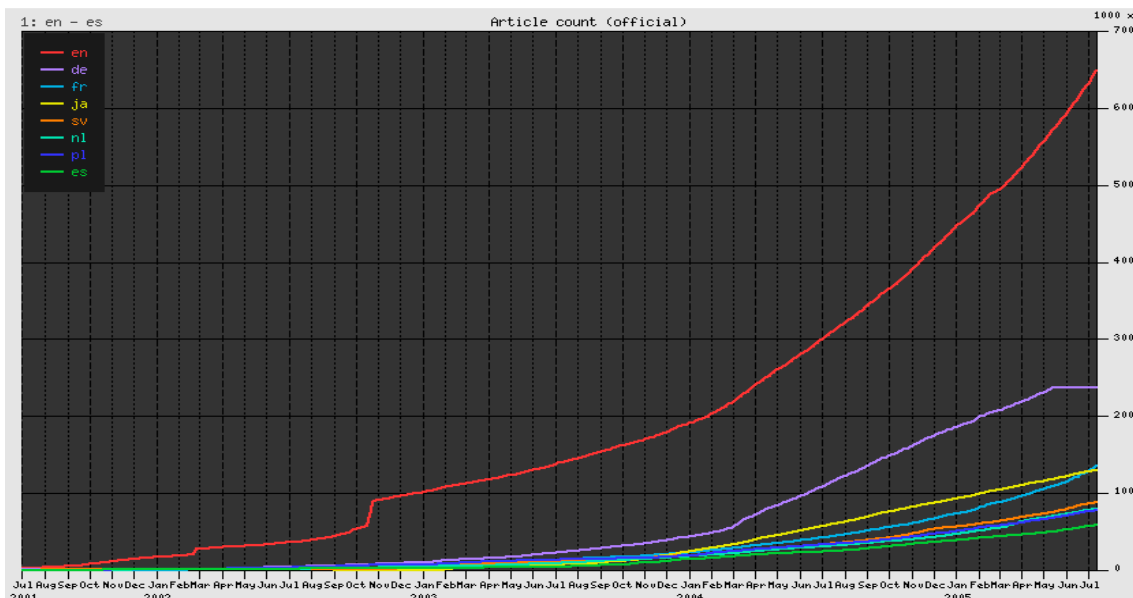
Tabelle 12 Anzahl der Artikel je Sprache in der Wikipedia. Quelle: Wikipedia [3], 2005.

<sup>12</sup> Zu dieser Zeit noch nicht im Internet im heutigen Sinne online verfügbar.

Grafisch lässt sich die Gewichtung leichter erfassen:



Obwohl die japanische Variante der Wikipedia nur zwei Monate nach der deutschen Wikipedia online ging, liegt sie mit etwas weniger als der Hälfte der Anzahl der deutschen Artikel deutlich zurück. Und das, obwohl es mehr japanische Internetnutzer gibt als deutsche (vgl. Global Reach [2], 2004). Deutsche scheinen also mehr Inhalte einzuspeisen, in schaffender Hinsicht also aktiver zu sein als Japaner. Anhand der historischen Entwicklung lässt sich dies auch nachweisen:



*Abbildung 6 Entwicklung der Artikelanzahl in der Wikipedia. Es ist deutlich erkennbar, dass Artikel auf Englisch und Deutsch mehr als andere Sprachen ansteigen. Quelle: Wikipedia [4], 2005*

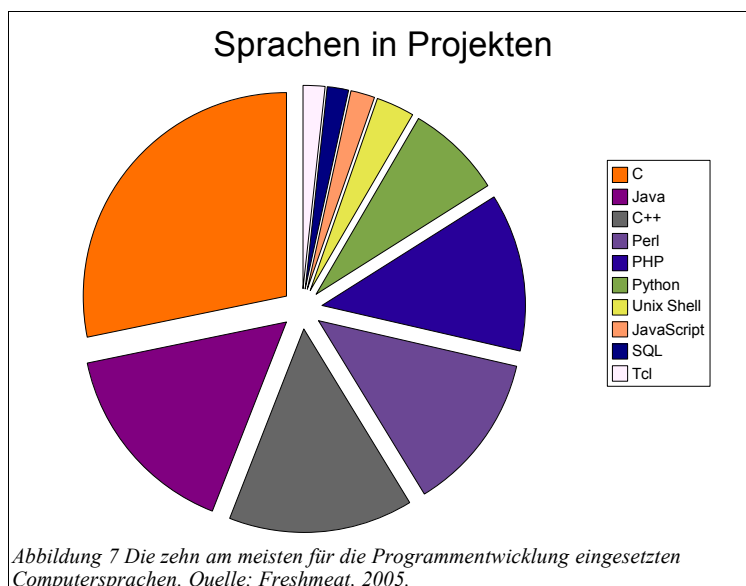
Doch wozu wird die Wikipedia hier aufgeführt bzw. was ist ihr Zusammenhang zur Sprachpolitik? Nun, ein unmittelbarer Zusammenhang besteht freilich nicht, schließlich wird die Wikipedia von keiner Regierung verwaltet oder aktiv gesteuert. Jedoch schafft eine Regierung mit einer guten Politik, die ihren Bürgern eine wirtschaftlich vorteilhafte Stellung beschert die Rahmenbedingungen dafür, dass ihre Bürger derart aktiv sein können.

Viel wichtiger sind allerdings die Folgeeffekte, die die freie Verfügbarkeit einer derart großen Enzyklopädie mit sich bringt: Internetnutzer, denen in ihrer Muttersprache kein geeignetes Nachschlagewerk zur Verfügung steht, werden sich denen zuwenden, die sie verstehen können, die also in einer ihnen bekannten Fremdsprache existieren. Bei einem Projekt wie Wikipedia ist es durchaus wahrscheinlich, dass bei einer entsprechend großen Anzahl von Nutzer auch welche von ihnen selbst beitragen, auch wenn es nicht ihre eigene Muttersprache ist. Dies führt zu einer immer größeren Sammlung von Wissen in ein paar wenigen Sprachen, in diesem Fall Deutsch und Englisch, was wiederum im Laufe der Zeit Einfluss auf die Bedeutung der Sprachen sowohl innerhalb als auch außerhalb des Netzes haben kann.

### 5.7.3. Spezialfall Programmiersprachen

Bisher wurden Belege für eine vielsprachige Zukunft des Internets erbracht. Hier soll nun auch anhand des Spezialfalls der Programmiersprachen, mit denen Computerprogramme geschrieben oder HTML-Seiten beschrieben werden, gezeigt werden, dass sich in bestimmten Bereichen eine einzelne Sprache als de facto lingua franca etablieren konnte.

Die folgenden Daten wurden dem Softwareportal „Freshmeat“ (<http://www.freshmeat.net>) entnommen. Freshmeat ist ein stark frequentiertes Portal zur Ankündigung von neuen Programmen oder Programmversionen. Besonders zahlreich vertreten ist kostenlose und freie Software<sup>13</sup>, deren Entwicklung meist über das Internet koordiniert wird, also sozusagen „im Internet entsteht“.



<sup>13</sup> „Frei“ bedeutet, dass die Software meist im Quellcode verfügbar ist und unter Einhaltung bestimmter Bedingungen auch ohne Genehmigung verändert und verbreitet werden darf.

Die oben aufgeführten Programmiersprachen, die in ca. 91% (vgl. Freshmeat, 2005) der Fälle zur Programmentwicklung eingesetzt werden, verwenden alle englische Begriffe als Kommandos. So steht etwa „print“ („drucke“) in Java für „gib auf den Bildschirm aus“ oder „alert“ („alarmiere“) in JavaScript für „weise den Benutzer mit einer Dialogbox auf etwas hin“. Sprachen wie z. B. LOGO, die deutsche Begriffe wie „drucke“ verwendeten, sind nicht gebräuchlich. Demnach muss jeder Programmierer sich zumindest ein wenig ins Englische einarbeiten. Für die Softwareentwicklung hat sich diese Sprache als lingua franca also fest etabliert.

## 6. Politische Vorgaben und Streitthemen

Die Sprachpolitik von Staaten oder Staatenbünden scheint nach den bisherigen Erkenntnissen wenig Einfluss auf das Internet gehabt zu haben. Ereignisse in jüngster Zeit lassen aber auch eine andere Schlussfolgerung zu: Dass sich die Politik nicht ernsthaft dem neuen Medium Internet zugewandt hat. Das aber würde bedeuten, dass sich bei einer anderen, zielgerichteten Sprachpolitik die Sprachverteilung im Internet ganz anders als bisher prognostiziert entwickeln könnte. Gezeigt werden soll dies am Fall der Digitalisierung von Schriftgut:

Alles begann am 14.12.2004 mit der Pressemeldung der Suchmaschine Google, die Bestände mehrerer großer angelsächsischer Bibliotheken digitalisieren und in ihren Index aufzunehmen (vgl. Braun, 2004). Für diesen Zweck freigegebene Bücher oder Bücher, deren Urheberrecht abgelaufen ist, sollen sogar vollständig online verfügbar sein.

Im März 2005 begann Frankreich, sich zunehmend politisch gegen dieses Projekt zu stemmen, da es eine „Furcht vor kultureller Dominanz der US-Weltsicht“ empfindet (Kuri [1], 2005). Zuvor wurden Frankreichs Bemühungen salopp<sup>14</sup> als ineffektiv bezeichnet und dass sie nur ein „verzweifelter Versuch“ (Rötzer, 1997) wären, „Gewordenes einzufrieren“. Frankreich schien damals allein zu stehen.

Dieses Mal aber ist die Lage anders: Mit großen Tempo gewann Frankreich an Unterstützung. So hatten kurz darauf 19 europäische Bibliotheken gefordert, durch die Vernetzung nationaler Projekte eine europäische digitale Bibliothek als Gegenpol zu den Plänen von Google zu schaffen (vgl. Kuri [2], 2005). Dieser Forderung wurde anschließend durch einen gemeinsamen Aufruf von den sechs EU-Staaten Deutschland, Frankreich, Italien, Spanien, Polen und Ungarn an den EU-Ratspräsidenten Nachdruck verliehen.

Bereits am 3. Mai befürwortete die EU offiziell die Errichtung einer europäischen digitalen Bibliothek (vgl. Klein, 2005). Wie ernst dieses Thema genommen wird, lässt sich aus der Äußerung von EU-Ratspräsident Juncker schließen: „Wir müssen handeln“. Obwohl ihr keine offizielle Anfrage vorliegt, sagt auch die Schweizerische Landesbibliothek ihre Unterstützung zu.

Bereits im April, wird am 31. August gemeldet, wurde vom deutsch-französischen Ministerrat in Paris die Entwicklung einer eigenen Suchmaschine beschlossen (vgl. Ziegler [1], 2005) – ein noch deutlicherer

---

<sup>14</sup> Dieser Eindruck entsteht insbesondere durch die Stelle „... und eben hat Staatspräsident Jacques Chirac eine neue Gefahrenquelle verdeckt, das Internet, in dem noch wenig Französisch vertreten sei.“ (Rötzer, 1997)

Konfrontationskurs zu Google als zuvor. Es bleibt abzuwarten, ob die am 18. Oktober gestartete Digitalisierung europäischer Bücher (vgl. Bleich, 2005) diesem etwas Wind aus den Segeln nehmen kann.

Google selbst sieht sich auch zunehmenden Druck aus der Wirtschaft ausgesetzt. Zum Einen bildeten Konkurrent „Yahoo“ (<http://www.yahoo.de>) und das „Internet Archive“ (<http://www.archive.org>) eine Allianz namens „Open Content Alliance“ (vgl. Behr, 2005). Zum Anderen wird Google von immer mehr Seiten rechtlich unter Druck gesetzt (vgl. Ziegler [2], 2005; vgl. Wilkens [1 & 2], 2005).

Besonders hervorzuheben ist an der obigen Entwicklung, dass ein Projekt eines einzelnen amerikanischen Unternehmens in der Lage war, nicht nur die Aufmerksamkeit der europäischen Politik auf sich zu lenken, sondern sogar eine massive, konzentrierte Aktion zur Folge zu haben. Die zugrunde liegende Sorge, dass angelsächsisches Kulturgut durch seine bloße Präsenz im Netz eine Verdrängung anderen Kulturguts zur Folge haben könnte, erhärtet die zuvor aufgestellte These, dass auch ein Projekt wie Wikipedia politische Wirkung entfalten kann und deshalb die Sprachpolitik beeinflusst. Dann könnten konzentrierte politische Maßnahmen aber auch Einfluss auf das Netz nehmen, da nun ein Anknüpfungspunkt gefunden ist.

## 7. Fazit

Es ist schwierig, ein generelles Urteil über die Sprachpolitik im Internet zu ziehen. Zu unsicher ist die Datenbasis, zu viele widersprüchliche Entwicklungen lassen sich nachweisen.

Allgemein scheint jedoch ein leichter Vorteil für die Zukunft der Vielsprachigkeit im Netz zu bestehen, sofern man von speziellen Gebieten wie dem angeführten Fall der Programmiersprachen absieht. Selbst für die Wissenschaft ist eine Aussage wie „Englisch ist die lingua franca“ als verfrüht anzusehen: Zwar schreiben „über zwei Drittel der Wissenschaftler“ auf Englisch (Rötzer, 1997), doch sinkt der Anteil der Muttersprachler (vgl. Rötzer, 1997). Die Zahl der Zweitsprachler steigt immer weiter, und wie an den Global-Reach-Statistiken ablesbar ist, gibt es noch viel unausgeschöpftes Nutzerpotenzial im Internet, beispielsweise für Chinesisch. Mit der zunehmenden Anzahl der so vernetzten Personen würde gewiss auch die Anzahl der publizierten, für jeden erreichbaren chinesischen Forschungsergebnisse zunehmen und somit die relative Bedeutung von Englisch etwas zurückdrängen. Allerdings wird Englisch auf absehbare Zeit, also für einen Zeitraum von ca. zehn bis zwanzig Jahren, seine Dominanz verteidigen, bevor sich deutliche Veränderungen erkennen lassen werden.

Interessant wird auch die Zukunft von Sprachen wie Esperanto sein, die gerade dazu geschaffen wurden, eine globale Kommunikationssprache zu werden. Ob ihnen Erfolg beschieden sein wird, lässt sich bisher nicht sagen. Von einem absehbaren Misserfolg kann zum jetzigen Zeitpunkt jedoch noch lange nicht gesprochen werden: Eine einsprachige Zukunft, so sie denn existieren wird, muss nicht zwangsweise dem Englischen gehören (vgl. Korpela, 2003).

## 8. Literaturverzeichnis

Folgende Sekundärliteratur wurde verwendet:

[Amazon]. „Search Inside the Book“.

<<http://www.amazon.com/exec/obidos/tg/browse/-/10197021/104-8260283-6179139>> (27.10.2005)

Behr, Bernd. 03.10.2005. „Neue Allianz will Bücher digitalisieren“.

<<http://www.heise.de/newsticker/meldung/64525/>> (24.10.2005)

Bleich, Holger. 18.10.2005. „Europäischer Ableger von Google Print gestartet“.

<<http://www.heise.de/newsticker/meldung/65043/>> (24.10.2005)

Braun, Herbert. 14.12.2004. „Google nimmt Bestände von Universitätsbibliotheken in Index auf“.

<<http://www.heise.de/newsticker/meldung/54218/>> (24.10.2005)

[Freshmeat]. „Freshmeat“. <<http://freshmeat.net/>> (27.10.2005)

[Gigablast]. 2005. <<http://www.gigablast.com>> (26.10.2005 23:50 Uhr)

[Global Reach 1]. 30.03.2004. „Global Internet Statistics“.

<<http://www.glreach.com/globstats/>> (23.05.2005)

[Global Reach 2]. 30.03.2004. „Evolution Of Online Linguistic Populations“.

<<http://global-reach.biz/globstats/evol.html>> (23.05.2005)

[Google]. 2005. <<http://www.google.com>> (28.05.2005 21:20 Uhr)

Gordon, Raymond G., Jr. (ed). 2005. „Ethnologue language name index“

<[http://www.ethnologue.com/language\\_index.asp](http://www.ethnologue.com/language_index.asp)> (26.10.2005)

Hashimoto, Satoshi; EURAC. 2005. „Japan steckt in einer monolinguischen Sackgasse“.

<[http://www.eurac.edu/Research/haschimoto\\_de.htm](http://www.eurac.edu/Research/haschimoto_de.htm)> (23.05.2005)

Hauck, Peter. 23.06.2005. „Kryptologie und Datensicherheit“.

<<http://www-dm.informatik.uni-tuebingen.de/skripte/Kryptologie/Kryptologie.pdf>> (26.10.2005)

Horvath, John. 24.03.1997. „Die Lingua Franca des Netzes“.

<<http://www.heise.de/tp/r4/artikel/1/1150/1.html>> (23.05.2005)

[Kazehakase]. „風博士(Kazehakase)“.

japanisch: <<http://kazehakase.sourceforge.jp/ja/?date=200408>> (23.05.2005)

englisch: <<http://kazehakase.sourceforge.jp/?date=200408>> (23.05.2005)

Klein, Ursula. 03.05.2005. „Schweiz und EU für Einrichtung von europäischer digitaler Bibliothek“.

<[http://www.kleinreport.ch/print\\_meld.phtml?id=28062](http://www.kleinreport.ch/print_meld.phtml?id=28062)> (27.10.2005)

[Koinzidenzindex]. 30.08.2005. „Koinzidenzindex“.

<<http://de.wikipedia.org/wiki/Koinzidenzindex>> (26.10.2005)

Korpela, Jucca. 02.09.2003. „English – the universal language on the internet?“.

<<http://www.cs.tut.fi/~jkorpela/lingua-franca.html>> (23.05.2005)

Kuri, Jürgen [1]. 21.03.2005. „Gegenwind für Google aus Frankreich“.

<<http://www.heise.de/newsticker/meldung/57770/>> (24.10.2005)

- Kuri, Jürgen [2]. 28.04.2005. „Sechs EU-Staaten für elektronische Bibliothek“.  
 <<http://www.heise.de/newsticker/meldung/59118/>> (24.10.2005)
- Kuri, Jürgen [3]. 04.10.2005. „EU plant Retrodigitalisierung von Bibliotheksbeständen“.  
 <<http://www.heise.de/newsticker/meldung/64536/>> (24.10.2005)
- [Netz-Tipp]. 2000. „Sprachen im Internet: Das Internet spricht Englisch ... und neuerdings auch Deutsch“.  
 <<http://www.netz-tipp.de/sprachen.html>> (23.05.2005)
- [Ogwehoweh]. 12.12.2003. „Ohwejagehka: Ha`degaenage“.  
 <<http://www.ohwejagehka.com/index.html>> (23.05.2005)
- Rötzer, Florian. 09.04.1997. „Sprachliche Monokultur oder Sprachenvielfalt?“.  
 <<http://www.heise.de/tp/r4/html/result.xhtml?url=/tp/r4/artikel/1/1157/1.html>> (27.10.2005)
- Sullivan, Daniel. 27.09.2005. „End Of Size Wars? Google Says Most Comprehensive But Drops Home Page Count“.  
 <<http://searchenginewatch.com/searchday/article.php/3551586>> (26.10.2005)
- Treanor, Paul. „Die Zukunft der Sprache in Europa“.  
 <<http://www.heise.de/tp/r4/artikel/1/1257/1.html>> (23.05.2005)
- [Web Languages Hit Parade]. Juni 1997. „Web Languages Hit Parade“.  
 <<http://alis.isoc.org/palmares.en.html>> (26.10.2005)
- Wilkens, Andreas [1]. 21.09.2005. „Google wegen 'massiver Copyright-Verletzungen' verklagt“.  
 <<http://www.heise.de/newsticker/meldung/64121/>> (24.10.2005)
- Wilkens, Andreas[2]. 12.08.2005. „Google pausiert beim Einscannen urheberrechtlich geschützter Bücher“.  
 <<http://www.heise.de/newsticker/meldung/62770/>> (24.10.2005)
- [Wikipedia [1]]. 27.10.2005. „History of Wikipedia“.  
 <[http://en.wikipedia.org/wiki/History\\_of\\_Wikipedia](http://en.wikipedia.org/wiki/History_of_Wikipedia)> (27.10.2005)
- [Wikipedia [2]]. 25.10.2005. „Wikipedia:Willkommen“.  
 <<http://de.wikipedia.org/wiki/Wikipedia:Willkommen>> (27.10.2005)
- [Wikipedia [3]]. „Wikipedia“.  
 <<http://www.wikipedia.org/>> (27.10.2005)
- [Wikipedia [4]]. 13.07.2005. „Wikipedia Statistics – Plots – Articles Count“.  
 <<http://en.wikipedia.org/wikistats/EN/PlotsPngArticlesTotal.htm>> (27.10.2005)
- [Yahoo]. 2005. <<http://www.yahoo.com>> (26.10.2005 23:55 Uhr)
- [Zeitgeist]. 2001-2005. „Google Zeitgeist“.  
 <<http://www.google.com/press/zeitgeist.html>> (23.05.2005)
- Ziegler, Peter-Michael [1]. 31.08.2005. „Frankreich forciert Pläne für europäische Google-Konkurrenz“.  
 <<http://www.heise.de/newsticker/meldung/63429/>> (24.10.2005)
- Ziegler, Peter-Michael [2]. 20.10.2005. „US-Verlegerverband macht Front gegen Google Print“.  
 <<http://www.heise.de/newsticker/meldung/65170/>> (24.10.2005)

# Anhang A

## Abbildungsverzeichnis

1	Entwicklung der Sprachverteilung im Netz. Quelle: Netz-Tipp, 2000.....	10
2	Entwicklung der Internetnutzer, die nicht Englisch sprechen. Stand: 30.03.2004. Quelle: Global Reach [2], 2004.....	14
3	Das Änderungsprotokoll eines japanischen Internetbrowsers namens KazeHakase. Hier die englische Version. Quelle: Kazehakase, 2005.....	15
4	Das Änderungsprotokoll eines japanischen Internetbrowsers namens KazeHakase. Hier die japanische Originalversion. Quelle: Kazehakase, 2004.....	16
5	Sprachverteilung der zehn Sprachen mit den meisten Artikeln innerhalb der Wikipedia. Quelle: Wikipedia [3], 2005.....	19
6	Entwicklung der Artikelanzahl in der Wikipedia. Es ist deutlich erkennbar, dass Artikel auf Englisch und Deutsch mehr als andere Sprachen ansteigen. Quelle: Wikipedia [4], 2005.....	19
7	Die zehn am meisten für die Programmentwicklung eingesetzten Computersprachen. Quelle: Freshmeat, 2005.....	20

# Anhang B

## Tabellenverzeichnis

1	Unterschiedliche Strömungen in Europa.....	4
2	Unterschiedliche Strömungen in Europa - erweiterte Darstellung .....	4
3	Die unterschiedlichen Ansätze zur Sprachanalyse einer Internetseite mit ein paar Beispielfverfahren....	7
4	Sprachverteilung 1997. Quelle: Web Languages Hit Parade, 1997.....	10
5	Sprachverteilung Anfang 2001. Quelle: Netz-Tipp, 2000.....	11
6	Sprachverteilung im Juni 2001. Quelle: Netz-Tipp, 2000.....	11
7	Sprachverteilung im Januar 2002. Quelle: Netz-Tipp, 2000.....	11
8	Sprachverteilung 2002. Quelle: Netz-Tipp, 2000.....	12
9	Unterschiedliche Ergebnisse für die Suche nach häufigen deutschen Wörtern mittels Google.....	12
10	Sprachverteilung von PDF-Dateien im Netz. Stand 2002. Quelle: Netz-Tipp, 2000.....	13
11	Ausschöpfungsgrad potenzieller Internetnutzer verschiedener ausgewählter Sprachen. Stand 30.03.2004. Quelle: Global Reach [2], 2004.....	13
12	Anzahl der Artikel je Sprache in der Wikipedia. Quelle: Wikipedia [3], 2005.....	18