

Matias Juuma

**OASIS XLIFF-STANDARDI JA SISÄLLÖN
LOKALISOINTI**



JYVÄSKYLÄN YLIOPISTO
TIETOJENKÄSITTELYTIETEIDEN LAITOS
2013

TIIVISTELMÄ

Juuma, Matias

OASIS XLIFF-standardi ja sisällön lokalisointi

Jyväskylä: Jyväskylän yliopisto, 2013, 30 s.

Tietojärjestelmätiede, Kandidaatin tutkielma

Ohjaaja: Jauhiainen, Eliisa

Tässä kandidaatintutkielmassa kuvataan sisällön lokalisointia sekä siihen liittyviä käsitteitä ja etuja. Sisällön lokalisointia esitellään eri näkökulmista ja sen hyötyjä ja ominaisuuksia on pyritty selventämään. Sisällön lokalisointia kuvataan niin organisaation näkökulmasta kuin myös käyttäjän eli sisällön kuluttajan näkökulmasta. Erilaisia standardeja on myös esitelty, jotta saataisiin vertailukohtia tämän tutkielman toiseen osaan eli XLIFF-standardiin. XLIFF-standardista annetaan yleiskuva siitä millainen standardi on kysessä ja mitä sillä voidaan tehdä. XLIFF-standardin ominaisuuksia on havainnollistettu merkkauseimerkkien avulla.

Tutkielma tarjoaa loppupäätelmänä yleiskuvan sisällön lokalisoinnin eduista ja siitä onko sisällön lokalisoinnista oikeasti mitään hyötyä organisaatiolle tai käyttäjälle. Lisäksi XLIFF-standardin sopivuutta sisällön lokalisoinnin standardiksi tullaan käsittelemään. Tutkielma kuvaa myös sisällön lokalisointia sisällön hallinnan kentässä eli sen suhdetta sisällön hallintaan.

Kysymyksiä: "Miksi sisältöjä lokalisoidaan?", "Mitä standardeja sisällön lokalisointiin on?" sekä "Mitkä ovat XLIFF-standardin ominaispiirteet?" "Onko XLIFF-standardi hyvä sisällön lokalisoinnin standardi?" käsitellään tässä tutkielmassa.

Asiasanat: XLIFF-standardi, sisällön lokalisointi, sisällön hallinta, asiakastytyväisyys, kääntäminen, käännosmuisti

ABSTRACT

Juuma, Matias

OASIS XLIFF-standard and localization of content

Jyväskylä: University of Jyväskylä, 2013, 30 p.

Information Systems Science, Bachelor's thesis

Supervisor: Jauhiainen, Eliisa

On this bachelor's thesis will be described content localization, and its related concepts and benefits. Content localization will be presented from different perspectives and the benefits and efforts are tried to clarify to the reader. Content localization is described from organizations perspective, but as well as (content) users' perspective. Different kinds of localization standards are presented in order to obtain reference points to this thesis's second part, XLIFF-standard. From XLIFF-standard there will be presented overview from XLIFF-standard and what can actually be done with it. XLIFF-standards features have been illustrated by examples.

As a conclusion thesis will give overview from benefits of content localization and does content localization actually give benefits for organizations or users and what are they. Also XLIFF-standards compatibility as a standard of content localization will be presented. Thesis also shows how content localization will indicate to content management.

Questions: "Why contents are localized? ", "What kind of standards there are for content localization?", and "What are XLIFF-standards basic-features?" will be handled.

Keywords: XLIFF-standard, localization, content management, customer satisfaction, translation, translation memory

KUVIOT

KUVIO 1 Sisällön hallinnan komponentit	8
KUVIO 2 TMX-esimerkkirakenne	13
KUVIO 3 TBX-perusrakenne.....	14
KUVIO 4 SRX-standardin perusrakenne	15
KUVIO 5 OLIF-standardin perusrakenne.....	16
KUVIO 6 xml:tm-esimerkki	17
KUVIO 7 XLIFF-standardin perusrakenne.....	20
KUVIO 8 HTML-esimerkkikuva	21
KUVIO 9 XLIFF-esimerkkikuva -elementin lokalisoinnista	22
KUVIO 10 SVG-esimerkkikuva	22
KUVIO 11 HTML-koodi SVG-kuvan tekstistä.....	22
KUVIO 12 XLIFF-käännös HTML-tiedoston SVG-kuvasta.....	23
KUVIO 13 XLIFF-tiedosto kohdekielille käännettynä.....	23
KUVIO 14 SVG-kuva käännettynä.....	23

SISÄLLYS

TIIVISTELMÄ	2
ABSTRACT	3
KUVIOT	4
SISÄLLYS.....	5
1 JOHDANTO.....	6
2 SISÄLLÖN LOKALISOINTI.....	8
2.1 Sisällön lokalisointi sisällön hallinnassa	8
2.2 Vaatimusanalyysi sisällön lokalisoinnissa	10
2.3 Lokalisointityökalut	11
2.4 Lokalisoinnin standardit.....	12
2.4.1 Translation Memory eXchange (TMX).....	12
2.4.2 TermBase eXchange (TBX).....	13
2.4.3 Universal Terminology eXchange (UTX).....	14
2.4.4 Segmentation Rules eXchange (SRX)	14
2.4.5 Open Lexicon Interchange Format (OLIF)	15
2.4.6 Xml:tm.....	16
2.4.7 Global Information Management Metrics eXchange (GMX).....	17
3 OASIS XLIFF-STANDARDI	19
3.1 OASIS XLIFF-standardina	19
3.2 XLIFF-standardin perusrakenne	20
3.3 XLIFF-standardi kuvan lokalisoisemisessa	21
3.4 Kokemukset XLIFF-standardista.....	23
4 YHTEENVETO	25
LÄHTEET	28

1 JOHDANTO

Teknologian kehityksen myötä ihmiset ovat päässeet ennen näkemättömän lähelle toisiaan. Ihmisten on internetin avulla helppoa olla yhteydessä toisiinsa ja ylläpitää suhteitaan tuttaviansa ja ystäviinsä ympäri maailmaa. Näin ollen myös liiketoiminnat ja palvelut ovat huomanneet globalisaation mahdollisuudet myös verkon kautta ja siksi yritykset ja eri organisaatiot tarjoavatkin palveluitaan kautta maailman. Varsinkin suurilla yrityksillä on intressejä laajentua ja tarjota palvelujensa käyttäjille ja asiakkaille mahdollisimman hyviä tuotteita ja palveluita.

Liiketoiminnassa tärkeänä osa-alueena olevaa asiakastyytyväisyyttä pyritään ylläpitämään ja kasvattamaan vaihtelevalla menestyksellä. Tärkeänä osa-alueena asiakastyytyväisyyden kasvattamisessa ja ylläpitämisessä on asiakkaiden ja käyttäjien tarpeisiin vastaaminen. Yksinkertaisesti sanottuna voisikin todeta asiakkaiden valitsevan sen tuotteen tai palvelun, joka parhaiten vastaa heidän tarpeitaan.

Sisällön lokalisoinnilla on asiakastyytyväisyyden parantamiseen paljon tarjottavaa. Monikansallinen suuri organisaatio voidaan mieltää kasvottomana ja ainoastaan voittoa tavoittelevana osakeenomitajien rahasampona. Käyttäjää tai asiakasta (sisällön kuluttajaa) tulisi pystyä palvelemaan niin hyvin että hänen tarpeensa tulee tyydytetyksi. Sisällön lokalisoinnilla onkin suuri merkitys tässä suhteessa, koska sillä voi olla suuri vaikutus sisällön kuluttajaan, millä kielellä hän sisältöä kuluttaa. On aivan eri asia vierailta englannin kielisellä sivulla, jonka sivuasettelu, kieli ja jopa värit ovat erilaisia kuin mihin esimerkiksi suomalainen käyttäjä on totunut. Tästä syystä sisällön lokalisoinnilla voidaan saavuttaa jopa selvää kilpailuetua kansainvälisillä markkinoilla. Miller (2006) toteaaakin että lokalisatio ei ole helppo työtehtävä, jonka voi vain liittää osaksi sisällön hallintajärjestelmää vaan se vaatii tarkkaa suunnittelua ja valmistelua riippuen kuitenkin organisaation tarpeista ja pyrkimyksistä globaalissa maailmassa. Tärkeintä kuitenkin on löytää yhteinen kieli sisällön kuluttajien ja organisaation välille.

Salo (2009) nostaa opinnäytetyössään toiseksi näkökohdaksi kehittyvien tai alemman kehityksen maiden internet-käyttäjien kasvava joukko.

Hän väittää että suurin osa internetin käyttäjistä käyttää internetissä kielenään englantia, mutta sisällön lokalisoinnin kannalta mielenkiintoista hänen väittämässään on että vuosien 2000 ja 2008 välisenä aikana arabian-, kiinan-, ja venäjänkielisten käyttäjien määrä kasvoi eniten suhteessa muuhun maailmaan. On siis selvää että tulevaisuudessa tarve sisällön lokalisoimiselle ja sen osaajille tulee kasvamaan entisestään.

Anastasioun (2010) mukaan sisällön lokaliointi ei ole pelkästään lähdekieleltä kääntämistä kohdekielelle. Toisin sanoen sisällön lokalisointi ei tarkoita esimerkiksi ainoastaan englannin kielisen datan kääntämistä venäjäksi. Sisällön lokalisointi kohdistuu myös kuvien, audion, videoiden, värien ja sisällön muokkaamiseen, jotta ne vastaisivat sisällön kuluttajan tarpeita. Yleisesti sisällön lokalisointi tarkoittaa kokonaisuutena sisällön muokkaamista, johonkin toisen käyttäjäryhmän tarpeisiin. Sisällön lokalisointia ei siis tulisi mieltää pelkkänä käännoistyönä vaan sitä tulisi ajatella laaja-alaisemmin. Vaikka käännoistyö kieleltä toiselle on kuitenkin merkittävä osa sisällön lokalisointia.

On olemassa erilaisia standardeja, jotka pyrkivät helpottamaan käännoistyötä. Standardit vaikuttavat erityisesti käännoismuistien eli käännettävien tekstien tietokantojen toimintaan. Toisin sanoen standardi vaikuttaa siihen miten kieltä tullaan kääntämään kieleltä toiselle ja miten se on tallennettuna käännoismuistissa.

Standardeja, jotka liittyvät sisällön lokalisointiin on maailmassa useita kuten esimerkiksi TMX ja SRX. Standardeja hallinnoivat isot voittoa tavoittelemattomat järjestöt kuten esimerkiksi OASIS (Organization for the Advancement of Structured Information Standards). Tässä tutkielmassa huomio on XLIFF-standardissa, koska se pyrkii tarjoamaan ainoastaan yhtä liittymää monien eri käännoismuistien välille. Muitakin standardeja esitellään lyhyesti luvussa kaksi.

XLIFF-standardi on lyhenne sanoista XML Localization Interchange File Format (XLIFF specification, 2008). XML lyhenne tulee puolestaan englannin kielen sanoista Extensible Markup Language (Bray, T. ym., 2006). Honkarannan & Jauhiaiaisen (2006) mukaan XML on kehitetty alunperin elektroniseen julkaisuun ja tekniseen dokumentointiin, jotta datan uudelleen käyttöä voitaisiin hallita. Organisaatiot käyttävät XML:ää järjestelmien integroinnissa, elektronisessa liiketoiminnassa sekä verkkopalveluissa.

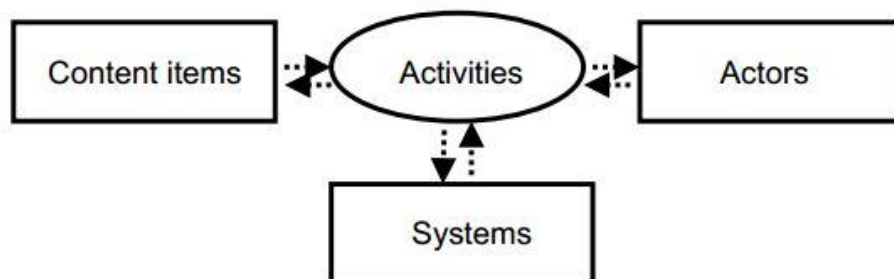
Tässä tutkielmassa kuvataan sisällön lokalisointia sekä siihen liittyvää käsitteistöä. Tutkielmassa annetaan vastauksia kysymyksiin "Miksi sisältöjä lokalisoidaan?", "Mitä standardeja sisällön lokalisointiin on?", "Mitkä ovat XLIFF-standardin ominaispiirteet?" sekä "Onko XLIFF-standardi hyvä sisällön lokalisoimisen standardi?". Sisällöllä tarkoitetaan tässä tutkielmassa tekstiä, joka on luettavissa eli on luonnollista kieltä. Esimerkiksi käyttöohjeet ovat tämän tyyppistä sisältöä. Tämä siksi, koska sisällön lokalisoimiseen liittyvät standardit keskittyvät erityisesti kielen kääntämiseen. Luvussa kolme kuvataan taas XLIFF-standardia. Tässä luvussa pyritään kertomaan yleisellä tasolla millainen XLIFF-standardi on. XLIFF-standardin perustoiminnallisuuksia tullaan avaamaan sekä sitä mikä XLIFF-standardi on.

2 SISÄLLÖN LOKALISOINTI

Tässä luvussa tullaan käsittelemään sisällön lokalisointia. Luvussa esitetään peruskäsitteet sisällön lokalisointiin kuten käännosmuisti, sisällön lokalisoinnin standardit ja käännohjelmat. Lisäksi luku antaa vastauksen sisällön lokalisoinnin suhteesta sisällön hallintaan sekä siihen miksi sisältöä tulisi lokalisoida organisaation tai sisällön kuluttajan näkökulmasta. Luvussa esitellään myös muita sisällön lokalisointiin liittyviä standardeja. Huomioitavaa kuitenkin on että tässä ei esitellä kaikkia sisällön lokalisoinnin standardeja vaan esitellyt standardit luovat pohjan vertailulle XLIFF-standardin kanssa.

2.1 Sisällön lokalisointi sisällön hallinnassa

Salminen (2003) on kuvannut sisällön hallinnan jakaantuvan kolmeen eri komponenttiin, jotka ovat vuorovaikuksessa keskenään *aktiiviteettien* (activities) kautta (kuvio 1). *Sisältöyksiköt* (content items) on ensimmäinen komponentti. *Toimijat* (actors) on toinen komponentti. *Järjestelmä* (systems) on kolmas komponentti.



KUVIO 1 Sisällön hallinnan komponentit (Salminen 2003)

Kuviossa 1 sisältö edustaa sisältöyksiköitä, joita tuotetaan sisällönhallinnan aktiiviteeteissa. Toimijat tuottavat sisältöä suorittamiensa

aktiviteettien kautta. Järjestelmä taas koostuu laitteista, standardeista ja ohjelmistoista, joita käytetään sisällön hallinnassa.

Kuviota 1 voidaan soveltaa sisällön lokalisoinnissa, jolloin se kuvaa sitä mitä toimijoita sisällön lokalisoinnissa on ja mitä yhteistä sisällön lokalisoinnilla ja tavallisella sisällön hallinnalla ympäristönä on. Sisältönä voitaisiin pitää käännettävää sisältöä kuten esim. käyttöohje. Sisältö voi olla kääntämätöntä tai käännettyä, mutta sisällön lokalisoinnilla on tarkoitus tuottaa käännettyä sisältöä. Keskeinen aktiviteetti on sisällön kääntäminen ja se tapahtuu kääntäjien eli sisällön tuottajien toimesta. Kääntämisestä tulee kuitenkin sisällön lokalisointia kun mukaan otetaan kolmas komponentti eli järjestelmät. Järjestelmä koostuu käännöstyökaluista, XML-kielen standardeista sekä lokalisoinnin standardeista.

Sisältö ei itsessään ole ainoa asia, jota voidaan lokalisoida, vaan myös metadataa voidaan lokalisoida. Tämä helpottaa käyttäjälle näytettävän sisällön lokalisointia kun metatiedot ovat käännettynä omalle kielelle. Metadata tarkoittaa dataa datasta eli dataa, joka kuvailee tiedoston tai kohteen sisältöä. (Anastasiou, 2010). Yleisesti sisällön lokalisointi tarkoittaa sisällön ja esitettävän datan kääntämistä johonkin toiseen kulttuurin tai maahan vastaamaan kohteen kieltä, lakeja, mielityksiä ja tapoja. Sisällöllä tarkoitetaan tässä yhteydessä tekstiä, joka on käännettävä lähdekielestä kohdekieleen. Myös kuvat, jotka halutaan lokalisoida vastaamaan kohdekäyttäjryhmänsä kieltä ja kulttuuria on osa tämän tutkielman aihealuetta.

Riikosen (2006 s. 40) mukaan Yunker (2002) määrittää kirjassaan lokalisointitavat kolmeen eri vaihtoehtoon. Ensimmäisenä on *kattava lokalisointi*, joka tarkoittaa nimensä mukaisesti kokonaisvaltaista lokalisointia. Tässä tapauksessa kaikki käyttäjälle näkyvä data on lokalisoitu. Toisena vaihtoehtona on *lisäävä lokalisointi*. Tässä tapauksessa lokalisoitua sisältöä lisätään vain vähän kerrallaan. Huomioitavaa on kuitenkin että lopputulos voi olla sama kuin mitä kattavalla lokalisoinnilla saavutetaan. Kolmantena on *Räätälöity lokalisointi*, jolla sisältö räätälöidään vastaamaan jokaista käyttäjäryhmää aina erikseen. Tämä tarkoittaa sivuston uudelleensuunnittelua, joka voi tulla tarpeeseen mikäli alkuperäistä sivua ei voida muokata vastaamaan kohdekäyttäjien tarpeita. Tässä tutkielmassa sisällön lokalisointiin paneudutaan kuitenkin kielen ja kulttuurin kontekstista.

Lokalisoitavana kohteena voi olla niinkin yksinkertainen asia kuin päivämäärän esittäminen. Suomessa on vakiintunut käyttöön päivämäärä asettelu "pp.kk.vvvv", kun taas esimerkiksi Yhdysvalloissa vakiintunut tapa on esittää päivämäärä muodossa "kk.pp.vvvv". Aikaa lokalisoidaan jo kellonaikojenkin puolesta, sillä esimerkiksi Yhdysvalloissa on jo neljä eri aikavyöhykettä (pl. Hawaii ja Alaska), jolloin myös mahdollisesti esitetty aika tulee lokalisoida (Riikonen. 2006.).

Toinen merkittävä osa-alue sisällön lokalisoinnissa on luonnollisesti kieli. Kielen kääntäminen voikin olla asia, joka ensimmäisenä tulee mieleen sisällön lokalisoinnista.

Kieltä voidaan kääntää mille kielelle tahansa, mutta sisällön lokalisoinnissa on huomioitavaa myös se mikä työkalu soveltuu parhaiten millekin kielelle. Esimerkiksi Japanin kielen suhteen UTX voisi olla erinomainen standardi lokalisoida sisältö, sillä se on kehitetty Japanissa paikallisen kielen tarpeet huomioiden.

Syitä sisältöjen lokalisointiin on paljon, mutta suurimpana syynä voidaan pitää loppukäyttäjän eli sisällön kuluttajan tarpeita ja tuottaa mahdollisimman hyvä käyttäjäkokemus. Esselink (2000) määrittelee lokalisoinnin yhdeksi mahdolliseksi syyksi jo pelkästään sen että joissakin maissa laki vaatii tuotteen tai palvelun tuotettavan natiivilla kielellä. Lakien noudattamisen ohella hän määrittelee paikallisten markkinoiden tarpeisiin vastaamisen. Käytännössä tämä tarkoittaa sitä että yritys pyrkii vastaamaan käyttäjiensä tarpeisiin mahdollisimman hyvin. Tämä siksi, jotta käyttäjät kokisivat tuotetun sovelluksen tai palvelun mahdollisimman hyväksi ja näin ollen valitsisivat tämän palvelun tai sovelluksen omakseen.

2.2 Vaatimusanalyysi sisällön lokalisoinnissa

Kun lokalisoitua sisältöä (esim. verkkosivu tai käyttöohje) tuotetaan tulisi huomioida myös kenelle sisältöä ollaan tekemässä ja mitä sisältö on. Vaatimusanalyysillä voidaan selvittää mitä asiakas-, eli sisällön kuluttaja, oikeasti haluaa ja mitä sisällön lokalisoinnissa tulisi tapahtua. Mäki (2008) kuvaa pro gradu-tutkielmassaan vaatimusanalyysin aloitettavan tunnistamalla käyttäjäryhmä sekä sovellukseen liittyvät tavoitteet. Vaatimusanalyysiä vaikeuttaa se, että analyysiä laadittaessa ei usein tunnisteta tulevaa loppukäyttäjää, mikä tekee vaatimusanalyysin tekemisestä haasteellisen. Sisällön lokalisointia voi usein helpottaa se, että käyttäjäryhmä on selvillä ennen kuin käännoistyötä aletaan tehdä. Esimerkiksi jos lokalisointi tapahtuu ruotsin kieleen niin silloin käyttäjienkin voidaan olettaa olevan ruotsalaisia. Näin ollen tuleva käyttäjäryhmä on tunnistettu ja siksi voidaan keskittyä hyvän käyttäjäkokemuksen toteuttamiseen luomalla lokalisoitu versio toiseen kulttuuriin.

Toisin sanoen tilanne voidaan kärjistää siten ettei venäläiseen kulttuuriin haluta siirtää muslimimaihin tuotettua sisältöä. Tällainen ratkaisu ei siis palvele käyttäjää parhaalla mahdollisella tavalla. Mikäli sisällön tuottamiseen liittyy liiketoimintaa on kilpailun kannalta ensiarvoisen tärkeää huomioida asiakas. Asiakas maksaa siitä että saa jonkin palvelun tai tuotteen joka vastaa hänen tarpeitaan. Näin ollen yritys, joka ottaa asiakastytyväisyyden liiketoimintansa keskiöön osana liiketoimintafilosofiaansa säilyttää jo hankitut asiakkaansa (Dunham, 1991). Tätä voidaan pitää keinona pitkäaikaisten asiakassuhteiden luomiseen. Sisällön lokalisointi palveleekin käyttäjän tarpeita ja auttaa omalta osaltaan asiakastytyväisyyttä edistävää tietojärjestelmäkehitystä.

2.3 Lokalisointityökalut

Anastasioun (2010) mukaan XLIFF-standardia tukevat ainakin seuraavat käännöstyökalut: Swordfish, XTM, SDL TRADOS, Alchemy Catalyst ja memoQ. Esimerkiksi Swordfishillä voidaan kääntää tiedostoja XLIFF-formaattiin, tai toisin päin XLIFF-formaatista oikeaan tiedostoon. XLIFF-tiedostoja voidaan myös kääntää Sworfishin käyttöliittymän kautta, jolloin ei tarvitse kirjoittaa suoraan merkintäkielellä. Swordfishissä tavallisen tiedoston (esim. word-tiedoston) muuttaminen XLIFF-formaattiin tapahtuu siten, että lokalisointityökalu kulkee vaihe kerrallaan aloittaen kysymällä XLIFF-tiedoston nimeä sekä käännettävän tiedoston nimeä tai polkua. Tarkoituksena on siis tuottaa XLIFF-tiedosto. Seuraavassa vaiheessa ohjelma kysyy lähde- ja kohdekieltä. Käyttäjä voi myös valita oman käännösmuistinsa, jonka pohjalta käännöstyö voidaan toteuttaa. Lisäksi käyttäjä voi valita SRX-tiedoston, jota voidaan käyttää käännösprosessissa. Lopuksi ohjelma generoi valmiiksi käännetyn XLIFF-tiedoston asetusten ja valitun käännösmuistin perusteella (Swordfish III User guide. 2007). Tämän jälkeen on toki järkevää käydä läpi koneen kääntämä teksti ohjelman avulla ja kääntää XLIFF-tiedosto mahdollisimman tarkasti. Kun tämä on tehty voidaan generoida XLIFF-tiedosto takaisin alkuperäiseen formaattiinsa, jolloin käännöstyö on valmis.

Riikonen (2008) määrittelee vielä pro gradu-tutkielmassaan käännöstyökalulle suositelluiksi ominaisuudeksi sen että käännöstyökalu tallentaa johonkin tietokantaan jo käännettyt sanat, joita se sitten voi ehdottaa käännöstyön aikana mahdolliseksi käännökseksi.

Kulttuurikohtaiset erot voivat myös olla sisällön lokalisoinnin tarpeen määrittäviä tekijöitä. Esimerkiksi toisessa kulttuurissa voidaan olla totuttu siihen että, jokin tietty toiminnallisuus löytyy, tietystä kohtaa sivustoa. Näin ollen hyvän käyttäjäkokemuksen tuottamiseksi onkin järkevää sijoittaa näitä painikkeita kulttuurin omaksumiin paikkoihin.

Edellä mainituilla asioilla pyritään tuottamaan sisältöä, joka olisi mahdollisimman relevanttia kohderyhmälleen. Gommans, Krishnan ja Scheffold määrittelevät (2001) että web-sivusto on suunniteltava kohderyhmä huomioiden ja sisällön tulisi olla sellaista että se vastaa kohderyhmänsä tottumuksia ja mieltymyksiä. Huomioitavaa on myös että hyvän käyttäjäkokemuksen aikaansaamiseksi liiat toiminnallisuudet voidaan kokea häiritsevänä mikä taas laskee käyttäjien tyytyväisyyttä sovellusta, tuotetta tai palvelua kohtaan. Gommans, Krishnan, Scheffold toteavatkin (2001) että sisällön kuluttajan tulee kyetä navigoimaan helposti ja ilman suuria ongelmia, jotta käyttäjätyytyväisyys ei laskisi ja siirtyisi käyttämään jotain toista vaihtoehtoista sovellusta tai palvelua.

2.4 Lokalisoinnin standardit

Sisällön lokalisointiin liittyy läheisesti käännösmuisti. Riikonen (2006) määrittelee käännösmuistin tietokannaksi, jossa on valmiiksi käännettynä lokalisoitavat termit ja lauseet. Lisäksi käännösmuistin avulla voidaan helposti tarkastaa uutta sisältöä tuottaessa onko käännettävä termi tai lause jo valmiiksi käännettynä. Tämä helpottaa uuden sisällön tuottamista merkittävästi, sillä tällöin ei tarvita kääntäjän apua kuin tapauksessa, jossa käännettävä kohde on entuudestaan tuntematon eli sitä ei siis löydy käännösmuistista.

Anastasioun (2010) mukaan sisällön lokalisoinnin standardeille on luotu vuonna 2006 Multilingual Information Framework (MLIF). MLIF on luotu yleiseksi alustaksi erilaisille sisällön lokalisoinnin tekniikoille, jolla voidaan hallita näillä eri standardeilla tai tekniikoilla luotuja sisältöjä.

Seuraavaksi esiteltyt tekniikat ja standardit sisällön lokalisointiin ovat hyvin lähellä toisiaan, mutta niissä on kuitenkin joitain pieniä eroja. Pääsääntöisesti käännösmuistit on tallennettu näihin XML-kieleen perustuviin standardeihin (Hudik, T. Ruopp, A. 2011)

Lokalisoinnin standardit on kehitetty käännöstyökaluja varten. Standardit mahdollistavat koneavusteisen kääntämistyön siten että kääntäjä voi keskittyä ainoastaan käännöstyöhön sen sijaan että kiinnittäisi huomiota dokumentin muotoiluun ja näin ollen myös koodiin. Kun dokumentti on muunnettu standardin mukaiseksi on helppoa käännösohjelman avulla löytää käännettävät kohdat.

2.4.1 Translation Memory eXchange (TMX)

TMX (Translation Memory eXchange) on lokalisointitekniikka johon XLIFF-standardi joiltakin osin pohjautuu. XLIFF on standardina kuitenkin sen verran erilainen että, sille on ollut tarpeellista tehdä oma spesifikaationsa. TMX-spesifikaation viimeisin versio on vuodelta 2005.

TMX-tekniikka koostuu kahdesta kerroksesta tai dokumentista. Ensimmäisenä määritellään minkälainen tiedosto on kyseessä, kuten esimerkiksi fontit ja muotoilut. Toisessa osiossa taas on itse dokumentin sisältö eli lokalisointiin liittyvä data (Hudik, T. Ruopp, A. 2011).

TMX-tekniikka on Rayan (2005) mukaan myyjäneutraali eli voittoa tavoittelematon lokalisointistandardi, joka on luotu käännösmuistin hallintaan ja käyttöön. Toisin sanoen tällä standardilla on mahdollista vaihtaa lokalisoitavaa dataa käännösmuistin ja käännöstyökalujen välillä. TMX tiedoston rakenne selviää parhaiten kuvioista 2, jossa on määriteltynä käännös englannin kielestä ranskan kieleen sekä sen muotoilua.

```

<?xml version="1.0" ?>
<tmx version="1.4">
  <header creationtool="XYZTool" creationtoolversion="1.01-023"
    datatype="PlainText" segtype="sentence"
    adminlang="en-us" srclang="EN"
    o-tmf="ABCTransMem">
  </header>
  <body>
    <tu>
      <tuv xml:lang="EN">
        <seg>Text in <bpt i="1">&lt;B</bpt>bold<ept i="1">&lt;/B</ept>.</seg>
      </tuv>
      <tuv xml:lang="FR">
        <seg>Texte en <bpt i="1">&lt;B</bpt>gras<ept i="1">&lt;/B</ept>.</seg>
      </tuv>
    </tu>
  </body>
</tmx>

```

KUVIO 2 TMX.esimerkkirakenne (Opentag.com)

Kuviossa on <tuv>-elementissä määriteltynä kieli, jota tagin sisällä käytetään. <seg>-elementissä on käännettävä teksti eli tässä tapauksessa "text in". Myös muotoilu on määriteltynä tämän elementin sisään. Vastaavasti käänös tapahtuu samalla tavoin kuin yllä paitsi <tuv>-elementin sisään on määritelty kieleksi Ranska ja <seg>-elementissä on käännettynä "texte en".

2.4.2 TermBase eXchange (TBX)

TBX on LISAn (The Localisation Industry Standards Association) hallinoina standardi. LISA on on Yhdysvaltalainen järjestö, jonka tavoitteena on tuottaa teollisuuden yhdenmukaisia standardeja toiminnan helpottamiseksi. TBX-standardilla tuotetaan ja haetaan vieraskielisiä termejä. TBX on XML-pohjainen standardi ja se toimii yhteenliittymänä terminologisen informaation tuottamiseen eri sisätlöjen välillä (LISA, 2008). TBX-standardin mukainen rakenne näyttää tältä (kuvio 3):

```

<?xml version='1.0'?> <!DOCTYPE martif SYSTEM "TBXcoreStructV02.dtd">
<martif type="TBX" xml:lang="en">
<martifHeader>
  <fileDesc>
    <sourceDesc>
      <p>From an Oracle corporation termbase</p>
    </sourceDesc>
  </fileDesc>
  <encodingDesc>
    <p type="XCSURI">http://www.lisa.org/fileadmin/standards/tbx/TBXXCSV02.XCS</p>
  </encodingDesc>
</martifHeader>
<text>
  <body>
    <termEntry id="eid-Oracle-67">
      <descrip type="subjectField">manufacturing</descrip>
      <descrip type="definition">A value between 0 and 1 used in ...</descrip>
      <langSet xml:lang="en">
        <tig>
          <term id="tid-Oracle-67-en1">alpha smoothing factor</term>
          <termNote type="partOfSpeech">noun</termNote>
        </tig>
      </langSet>
      <langSet xml:lang="hu">
        <tig>
          <term id="tid-Oracle-67-hu1">Alfa simitási tényező</term>
          <termNote type="partOfSpeech">noun</termNote>
        </tig>
      </langSet>
    </termEntry>
  </body>
</text>
</martif>

```

KUVIO 3 TBX-perusrakenne (LISA)

Kuviossa on käännettynä englannin kielen termi "alpha smoothing factor" unkariksi. Käännökset ovat <tig>-elementin sisällä. Tässä tapausessa käännöksiä on vain yksi kappale ja se on <term>-elementin sisällä.

2.4.3 Universal Terminology eXchange (UTX)

Okuran, Bondin, Yamamoton, Muratan, Uchimoton, Katon, Shimazun, ja Suzukin (2009) mukaan UTX-standardi on samankaltainen kuin TBX-standardi, mutta se on suunniteltu helpommin käytettäväksi kuin mitä käännöstyön ammattilaisille suunnattu TBX-standardi. TBX-standardin tavoin UTX-standardi on suunniteltu lähinnä terminologian lokalisointiin ja tarkoitettu lähinnä koneelliseen käännöstyöhön. UTX:stä on olemassa myös UTX-S (UTX-Simple) versio, jonka tavoitteena on tehdä kääntämisestä helpompaa ja yksinkertaisempaa. UTX-standardi on suunnattu erityisesti aasian kielen lokalisointiin.

2.4.4 Segmentation Rules eXchange (SRX)

Kuten TBX- ja TMX-standardi myös SRX-standardi on LISA:n hallinnoima. SRX-standardi on tarkoitettu ilmaiseen jakeluun eli, sillä ei tavoitella kaupallista

hyötyä vaan se on suunniteltu yhtenäistämään lokalistointitapoja. SRX-standardi on suunniteltu kuvaamaan sisällön asettelua siltä osin kuin käännösohjelmat sitä prosessoivat. SRX-standardi suunniteltiin, koska havaittiin että TMX-standardi ei kykene kattamaan kaikkien käännösohjelmien tarpeita. TMX-standardi jäsentää sisällön aina tiettyyn muotoon, jolloin kaikki käännösohjelmat eivät kykene käyttämään sisältöä. (LISA, 2008) SRX-standardin perusrakenne näyttää tältä:

```
<?xml version="1.0" encoding="UTF-8"?>
<srx xmlns="http://www.lisa.org/srx20" version="2.0">
  <header segmentsubflows="yes" cascade="no">
    <formathandle type="start" include="no"></formathandle>
    <formathandle type="end" include="yes"></formathandle>
    <formathandle type="isolated" include="no"></formathandle>
  </header>
  <body>
    <languageules>
      <languageule languageule="default">
        <rule break="no">
          <beforebreak>([A-Z]\.){2,}</beforebreak>
          <afterbreak>\s</afterbreak>
        </rule>
        <rule break="yes">
          <beforebreak>\.</beforebreak>
          <afterbreak>\s</afterbreak>
        </rule>
      </languageule>
    </languageules>
    <maprules>
      <languageule languageule="*" languageule="default"></languageulemap>
    </maprules>
  </body>
</srx>
```

KUVIO 4 SRX-standardin perusrakenne (opentag.com/okapi. 2011)

Esimerkissä (kuvio 4) on määriteltynä käännösohje, jonka mukaan käännöprosessi voidaan suorittaa. <languageules>-elementti määrittää säännöt joiden mukaan käännösprosessi tehdään. <languageule>-elementti taas määrittää itse säännön.

2.4.5 Open Lexicon Interchange Format (OLIF)

OLIF-standardi tarjoaa tavan jäsentää lokalisoitavaa dataa tekstissä. OLIF-standardi lokalisoii yksittäisiä termejä ja sanoja riippumatta käännöstyökalusta eli se kykenee toimimaan monien eri käännöstyökalujen kanssa. Alun perin OLIF-standardi kehitettiin juuri käännöstyökalujen väliseen kommunikointiin, mutta myöhemmin sitä on käytetty myös muussa lokalisointi työssä. (McCormick, Lieske & Culum 2004) OLIF-standardin perusrakenne näyttää seuraavalta:

```

<entry ConceptUserId="0731F16CCCD2D3119B4D">
  <mono>
    <keyDC>
      <canForm>table</canForm>
      <language>en</language>
      <ptOfSpeech>noun</ptOfSpeech>
      <subjField>general</subjField>
      <semReading>86</semReading>
    </keyDC>
  </monoDC>
  .....
</monoDC>
</mono>
</entry>

<entry ConceptUserId="0731F16CCCD2D3119B4D">
  <mono>
    <keyDC>
      <canForm>Tabelle</canForm>
      <language>de</language>
      <ptOfSpeech>noun</ptOfSpeech>
      <subjField>general</subjField>
      <semReading>86</semReading>
    </keyDC>
  </monoDC>
  .....
</monoDC>
</mono>
</entry>

```

KUVIO 5 OLIF-standardin perusrakenne (McCormick, Lieske & Culum 2004)

Kuviossa 5 on <keyDC>-elementin sisälle kerätty lähdekielen metatietoja. Olennaisin löytyy <canForm>-elementin sisältä eli käännettävä sana "table". Engalnnin kielisen osuuden alapuolella taas on vastaavat tiedot saksaksi.

2.4.6 Xml:tm

Valmistuttuaan xml:tm-standardi on ollut kokonaan uusi tapa varastoida käännösmuistiin ja käyttää käännösmuistia. xml:tm on täysin integroitu XML-kieleen ja se noudattaa XML:n syntaksia kuvatessaan käännösmuistin rakennetta. Xml:tm-standardi on täysin integroitu XML-kieleen. Se rakentuu osaltaan TMX 2.0 spesifikaatiosta, SRX-standardista sekä se on integroitu tiukasti vastaamaan XLIFF-standardin spesifikaatiota. Näin ollen xml:tm-standardia voitaisiin pitää eräänlaisena kombinaationa näistä kolmesta edellä mainitusta standardista. Xml:tm-standardin tarkoituksena on yksinkertaistaa käänносprosessia XLIFF-standardin formaattiin ja siten helpottaa kääntämistä. Kuviossa 6 on esiteltyä miten XML-dokumentti näkyy xml:tm nimiavaruudella. xml:tm:n tarkoituksena on kuvata mitkä elementit sisältävät ei käännettäväksi tarkoitettua dataa ja mitkä elementit taas sisältävät käännettäväksi tarkoitettua dataa (Zydroń, Raya & Bogacki 2007).


```

<text:p text:style-name="Text body">
  <tm:ta id="a1" name="text:index-qualifier" version="1.0">
    xml:tm description
  </tm:ta>
  <tm:te id="e1" tu="2" version="1.0">
    <tm:tu id="u1.1" ti="1" crc="3275b242" version="1.0";>
      xml:tm is a radical new approach
      <text:index>
        <tm:ti id="i1.1.1" name="text:name" crc="9114ce48" version="2.0">
          radical new approach
        </tm:ti>
      </text:index>
      to dealing with the problems of translation
      memory for XML documents by using XML syntax to embed memory
      directly into the XML documents themselves.
    </tm:tu>
    <tm:tu id="u1.2" crc="306bf701" version="1.0">
      It makes extensive use of XML namespace.
    </tm:tu>
  </tm:te>
</text:p>
<text:p text:style-name="Text body">
  <tm:te id="e2" tu="2" version="1.0">
    <tm:tu id="u2.1" crc="f8c012ff" version="1.0">
      The "tm" stands for "text memory".
    </tm:tu>
    <tm:tu id="u2.2" crc="270af770" version="1.0">
      There are two aspects to text memory:
    </tm:tu>
  </tm:te>
</text:p>
<text:ordered-list text:continue-numbering="false" text:style-name="L1">
<text:list-item>
<text:p text:style-name="P3">
  <tm:te id="e3" tu="1" version="1.0">
    <tm:tu id="u3.1" crc="851603a2" version="1.0">
      Author memory
    </tm:tu>
  </tm:te>
</text:p>
</text:list-item>
<text:list-item>
<text:p text:style-name="P3">
  <tm:te id="e4" tu="1" version="1.0">
    <tm:tu id="u4.1" crc="313af159" version="1.0">
      Translation memory
    </tm:tu>
  </tm:te>
</text:p>
</text:list-item>

```

KUVIO 6 xml:tm-esimerkki (Zydroń, Raya & Bogacki 2007)

<tm:ta>-elementillä merkataan otsikko, jonka alapuolella ovat elementit ovat tämän lapsielementtejä. <tm:te>-elementillä kuvataan elementit, jotka sisältävät PCDATA-tekstiä. <tm:tu>-elementti sisältää <tm:te>-elementin varsinaisen tekstin. <tm:ti>-elementillä taas saadaan haettua kaikki käännettävissä olevat attribuutit, jotka ovat <tm:tu>-elementin sisällä <tm:ti>-elementin lapsielementeiksi.

2.4.7 Global Information Management Metrics eXchange (GMX)

GMX-standardikokoelma koostuu kolmesta versiosta, jotka ovat Volume (-V), Complexity (-C), sekä Quality (-Q). GMX-Volume on tarkoitettu arvioimaan työmäärää tai räsytystä, joka lokalisoitavalla työllä on. GMX-Volumella pystyy arvioimaan siis tarkemmin lokalisoinnin työmäärää kuin mitä sanojen

lukumäärää laskemalla. Se tuottaa siis tarkempaa dataa lokalisoitavan sisällön laajuudesta. GMX-Complexity:ä taas kuvaa lokalisoitavan kohteen haastavuutta tai kompleksisuutta kuten nimikin jo antaa olettaa. GMX-Quality taas mittaa lokalisoitavan sisällön laatua. Näistä kolmesta vain GMX-V on tällä hetkellä käytössä eli GMX-C ja GMX-Q ovat joko hylättyjä kehitelmiä tai sitten niitä vielä kehitetään.. (LISA, GMX)

3 OASIS XLIFF-STANDARDI

Tässä luvussa kuvataan XLIFF-standardin syntyä ja sen historiaa. Luvussa selvitetään XLIFF-standardin standardointiprosessia. Lisäksi luku esittelee XLIFF:n perustoiminnallisuuksia. Luvun lopusta löytyy Anastasioun (2010) tutkimuksen tiivistelmä XLIFF:n käyttäjäkokemuksista sisällön lokalisoinnin parissa työskentelevien ja opiskelevien henkilöiden mukaan.

3.1 OASIS XLIFF-standardina

XLIFF on avoin standardi, jonka avulla voidaan vaihtaa lokalisoitua dataa (Lewis, D. ym., 2009). Standardi siis määrittelee sen kuinka lokalisoitavaksi tarkoitettua dataa tulisi kirjoittaa merkintäkielellä. Merkintä vaihtelee standardin mukaisesti ja jokaisella standardilla on omanlaisensa merkintätapa. XLIFF-standardin tarkoituksena on siis helpottaa sisällön lokalisointia, jotta dataa olisi mahdollisimman yksinkertaista kääntää kieleltä toiselle. Rayan (2004) mukaan tärkein syy miksi XLIFF-standardia kannattaa käyttää on se että, XLIFF-standardilla voidaan kääntää erilaisia tiedostomuotoja. Toinen syy käyttää XLIFF-standardia on se että se antaa kääntäjälle mahdollisuuden keskittyä käännöstyöhön ilman keskittymistä käännettävän sisällön asetteluun.

OASIS XLIFF-standardi on saanut nimensä luojiensa mukaan eli OASIS-järjestön mukaan. OASIS-järjestö koostuu yksityisistä IT-alan yrityksistä ja toimijoista, joiden tavoitteena on kehittää IT-alan avointen yhtenäisten standardien ja kehitystä. XLIFF-standardoinnin luomisen ja jatkuvan kehittämisen lisäksi OASIS-järjestö toimii myös monissa muissa IT-alan standardointiin liittyvissä projekteissa (OASIS. 2012), mutta tämä tutkielma ei ota kantaa näihin muihin projekteihin ja standardeihin.

XLIFF-standardi on tapa, suorittaa sisällön lokalisointia. XLIFF-standardi on XML-pohjainen ja siten lokalisoitavaa dataa prosessoidaan XML-formaatissa. XLIFF-standardin mukainen xml-dokumentti tallennetaan .xlf-tiedostomuodossa.

XLIFF-standardi on kehitetty OASIS-järjestön toimesta, mutta idea XLIFF-standardin kehittämiseen lähti dublinilaisilta lokaalisiaation ja globaalisiaation ammattilaisilta. He kokivat että ei ole olemassa tarpeeksi tai olleenkaan standardia, joka pystyisi tuottamaan kustannustehokkaasti ja helposti lokalisoitavaa dataa. Asiantuntijat olivat yhteydessä Novelliin, Oracleen, Suniin, ja Berlitziin keskustellakseen mahdollisesta uudesta standardista. Työryhmä ei kuitenkaan päässyt muusta sopimukseen kuin että se pohjautuisi XML-kieleen. Tästä syystä otettiin yhteyttä eri standardointijärjestöihin kuten OASIS, W3C, ja LISA, joista OASIS valittiin standardin kehittäjäksi, koska se on niin läheisesti yhteydessä XML-kieleen. Lisäksi moni XLIFF-työryhmässä oleva organisaatio oli jo jäsenenä OASIS-järjestössä (Reynolds & Jewtushenko. 2005).

3.2 XLIFF-standardin perusrakenne

XLIFF-standardi toimii siten että standardin mukainen XML-merkkkaus upotetaan käännettävään XML-tiedostoon. Perusrakenteeltaan se on spesifikaationsa mukaan seuraavanlainen: (kuvio 7).

```
<xliff version='1.2'
  xmlns='urn:oasis:names:tc:xliff:document:1.2'>
<file original='hello.txt' source-language='en' target-language='fr'
  datatype='plaintext'>
  <body>
    <trans-unit id='hi'>
      <source>Hello world</source>
      <target>Bonjour le monde</target>
    <alt-trans>
      <target xml:lang='es'>Hola mundo</target>
    </alt-trans>
  </trans-unit>
</body>
</file>
</xliff>
```

KUVIO 7 XLIFF-standardin perusrakenne (XLIFF 1.2 specification)

xmlns-attribuutilla määritellään mitä nimiavaruutta käytetään. Nimiavaruudella tarkoitetaan nimikokoelmaa, joka on saatavissa jonkin tietyn URI:n avulla (Bray, T. Hollander, D. Layman, A. 1998). XLIFF-dokumentin yhteydessä käytetään nimiavaruutena seuraavaa: "urn:oasis:names:tc:xliff:document:1.2".

Perusrakenteesta on havaittavissa XLIFF:n versionumero, joka siis on uusin eli 1.2. <file>-elementti määrittelee metadataa itse tiedostosta, joka on tässä esimerkkitapauksessa tekstitiedosto "hello.txt". Lisäksi on määritelty lähdekieleksi englanti (source-language='en') sekä kohdekieleksi Ranska eli kieli, jolle tekstiä aiotaan lokalisoida (target-language='fr'). datatyypiksi on asetettu "plaintext", joka tarkoittaa sitä, että teksti on aivan tavallista, ymmärrettävää tekstiä.

Tiedoston <file> sisällä oleva rakenne <body> koostuu käännösyksiköistä tai -lohkoista <trans-unit>, jotka voidaan spesifioida käyttämällä id-attribuuttia esimerkkikuvion osoittamalla tavalla. Tässä tapauksessa lokalisointi on spesifioitu termillä "hi". Tämän alle on määritelty se mitä <trans-unit>-elementti kääntää. Kääntäminen tapahtuu näiden <trans-unit>-elementtien sisällä (Reynolds & Jewtushenko. 2005). Tässä tapauksessa lähteenä toimivasta tiedostosta (hello.txt) on poimittu englannin kielinen kohta "Hello world" ja sijoitettu se kohtaan "source". Lokalisointi tapahtuu siten, että <target>-elementin kohdalle sijoitetaan esimerkin mukaisesti ranskankielinen tervehdys: "Bonjour le monde". Huomioitavaa on, että yhtä hyvin tuossa kohtaa voisi lukea myös Suomen kielellä "Hei maailma" tai venäjäksi: "привет мир". Toisin sanoen, tässä kohtaa toteutetaan käännöstyö halutulle kielelle. Vaihtoehtoisesti voidaan lisäksi merkitä vaihtoehtoinen käännös elementillä <alt-trans>. Tämän sisälle voidaan esimerkin osoittamalla tavalla tehdä samainen käännös, mutta eri kielelle.

Käytännössä <target>-elementille sisälle on määriteltynä xml:lang attribuutin avulla käännettävä kieli ja sen jälkeen voidaan määrittää kuinka tuo "hello world" kääntyy kyseiselle kielelle. Tässäkin tapauksessa on huomioitava, että kieli voi olla mikä tahansa. Lopuksi elementit suljetaan kuten myös xliif-tiedosto. Yhteneväisyys xml-tiedoston syntaksiin on ilmeinen.

3.3 XLIFF-standardi kuvan lokalisoinnissa

XLIFF-standardilla on mahdollista myös lokalisoida kuvatekstejä ja siihen liittyvää metadataa kuten kuvatekstejä. Kuviossa 8 on esimerkkinä HTML-esimerkkikuva, jossa on <p>-elementin sisällä on kuvateksti "My picture" sekä -elementti, joka määrittelee kuvan. Lisäksi -elementin sisällä on alt-attribuutti, jolla määritellään vaihtoehtoinen kuvateksti.

```
<p>My picture,  
  
and there you have it.</p>
```

KUVIO 8 HTML-esimerkkikuva (Savourel, Schnabel, Jewtushenko & Domeny 2006)

Kuviossa 9 nähdään XLIFF-versio kuvasta, jossa kuvan metadataan on lisätty metadataa itse kuvasta. Tätä ominaisuutta voidaan siis käyttää web-sivujen kuvissa lisäämällä kohdekielen tekstiä kuvan metadataan alla olevaa esimerkkiä soveltaen

```

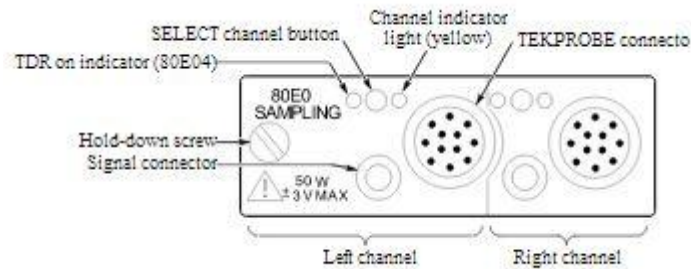
<trans-unit resname="p" id="d0e1">
<source xml:lang="EN">My picture,
  <ph id="d0e3" ctype="image" xhtml:src="mthood.jpg">
    <sub ctype="x-img-alt">This is a shot of Mount Hood</sub>
  </ph>
and there you have it.</source>

```

KUVIO 9 XLIFF-esimerkkikuva -elementin lokalisoinnista (Savourel, Schnabel, Jewtushenko & Domyeny 2006)

Käännöstyötä tehdessä kuvan lähdettä ei haluta kääntää, joten se on sijoitettu <ph>-elementin sisään. Kuvan vaihtoehtoinen teksti eli alt-attribuutti on tämän <ph>-elementin sisällä, joka näin ollen voidaan kääntää. Tämä on helpoin tapa suorittaa kääntäminen kuvien yhteydessä, sillä XLIFF-standardi mahdollistaa <ph>-elementin laajennettavuuden (Savourel, Schnabel, Jewtushenko, & Domyeny. 2006.).

Toinen mielenkiintoinen kuvan lokalisointi esimerkki koskee SVG-tiedostoja. XLIFF-standardin avulla voidaan lokalisoida SVG-tiedostomuodon kuvatekstejä seuraavasti: Alla oleva kuva esittää SVG kuvaa, joka sisältää tekstiä kuvan tulkitsemisen helpottamiseksi.



KUVIO 10 SVG-esimerkkikuva (Savourel, Schnabel, Jewtushenko & Domyeny 2006)

Kuviossa 10 on elektroninen laite ja siihen liittyvät tekstit, jotka selittävät laitteen toiminnallisuuksia.

Kuviossa 11 taas on esitetty koodi, joka määrittelee alkuperäisen kuvan tekstien sisällön.

```

<text id="XMLID_1_" transform="matrix(1 0 0 1 0 32.0503)">
  <tspan x="0" y="0" fill-rule="evenodd"
  clip-rule="evenodd" font-family=""TimesNewRomanPSMT"
  font-size="9.25">TDR on indicator (80E04)</tspan></text>

```

KUVIO 11 HTML-koodi SVG-kuvan tekstistä (Savourel, Schnabel, Jewtushenko & Domyeny 2006)

Kuviossa 11 on siis eriteltyä se miten teksti "TDR on indicator" näkyy kuvan päällä. koodissa on myös määriteltyä tekstin paikka kuvassa sekä sen muotoilu. XLIFF:n avulla kuvan teksti kääntyy näin:

```

<trans-unit id="A001-d0e499" resname="tspan">
<source>TDR on indicator (80E04)</source>
<target state="needs-translation">TDR on indicator (80E04)</target>
</trans-unit>

```

KUVIO 12 XLIFF-käännös HTML-tiedoston SVG-kuvasta (Savourel, Schnabel, Jewtushenko & Domyeny 2006)

Kuviossa 12 siis tunnistetaan kuvan käännettävä kohde. <source>-elementin sisällä on kohde, joka on käännettävä ja <target>-elementti taas on varattu kohdalle, joka on lokalisoitava. Kuviossa 13 taas esitetään se kuinka kuvien tekstit tämän jälkeen kääntyvät kohdekielille ja kuinka se voidaan muuttaa takaisin SVG-kuvaksi.

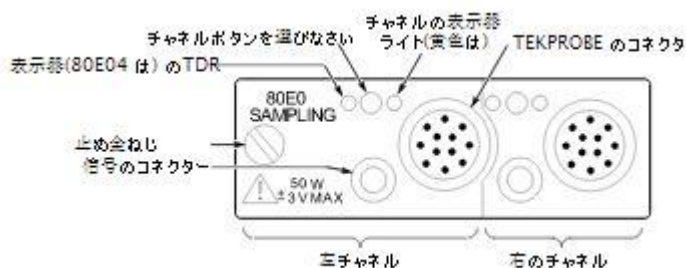
```

<text id="A001-XMLID_1_" transform="matrix(1 0 0 1 0 32.0503)">
<tspan id="A001-d0e499" x="0" y="0" fill-rule="evenodd"
clip-rule="evenodd" font-family="Arial Unicode MS"
font-size="9.25">表示器(80E04 は)のTDR</tspan></text>

```

KUVIO 13 XLIFF-tiedosto kohdekielille käännettynä (Savourel, Schnabel, Jewtushenko & Domyeny 2006)

Lopuksi vielä kuviossa 14 esitetään miltä käännetty SVG-kuva näyttää lokalisointiprosessin jälkeen.



KUVIO 14 SVG-kuva käännettynä (Savourel, Schnabel, Jewtushenko & Domyeny 2006)

3.4 Kokemukset XLIFF-standardista

Vaikka XLIFF kykenee käsittelemään paljon metadataa jakaa se silti mielipiteitä puolesta ja vastaan. Tästä syystä Anastasiou (2010) käsittelee artikkelissaan tutkimusta siitä miten opiskelijat ja ammattilaiset kokivat XLIFF-standardin. Tutkimukseen vastanneista suurin osa oli sisällön lokalisointiin liittyvien työkalujen tarjoajia sekä lokalisointipalvelun tarjoajia. Loput vastaajista olivat kääntäjiä, tutkijoita, konsultteja ja opiskelijoita.

Kyselyyn vastanneilta kysyttiin ovatko heidän teknologiansa ja käyttämänsä työkalut sisällön lokalisoinnissa yhteensopivia XLIFF-standardin kanssa. Vastaajista 33 % ilmoitti että heidän työkalunsa tukevat XLIFF-standardia kun taas 20 % vastaajista ilmoitti ettei heidän työkalunsa tue XLIFF-standardia. 20 % vastaajista totesi että osa tukee ja osa ei.

17 % vastanneista kertoi kuulleen XLIFF:stä ja olevansa kiinnostunut standardista, mutta ei syystä tai toisesta ole sitä käyttänyt. 3 % taas aikoi tulevaisuudessa tutustua XLIFF-standardiin kun taas 6 % vastaajista ei ollut koskaan kuullutkaan XLIFF-standardista. 1 % vastasi jotain muuta kuin edellä. Kolmantena kysymyksenä kysyttiin XML:n käytöstä XLIFF:ssä ja ennen kaikkea useiden eri nimiavaruuksien käytöstä. Tämä kysymys jakoi vastaajien mielipiteitä ja näin ollen tarkkana tuloksena tälle kysymykselle todettiin vain että osa vastaajista toivoi XLIFF:lle lisää laajennettavuutta kun taas osa vastaajista piti XLIFF-standardia muutenkin jo liian muokattavana standardina.

Tutkimuksessa kuitenkin todettiin että suurin osa vastaajista näki jokaiselle CAT-työkalulle eri nimiavaruuden käytön ongelmallisena, sillä se saattaa johtaa datan häviämiseen kun näitä CAT-työkaluja käytetään ristiin. Myös tukimuksen tekijät olivat yhtä mieltä tämän ongelman suhteen. Neljäntenä kysymyksenä vastaajilta kysyttiin sitä tulisiko XLIFF:llä olla enemmän synergiaa muihin sisällön lokalisoinnin standardeihin. Vastauksena käyttäjät totesivat yksiselitteisesti että kyllä, mutta vastaajat toivoivat synergiaa erityisesti standardeihin kuten TMX, IPS ja GMX-V. Viidentenä kysymyksenä kysyttiin ongelmista joita XLIFF:iä käyttäneet olivat kohdanneet. Vastauksista nousi esiin seuraavia ongelmia loppukäyttäjien parissa: Työkalut eivät tukeneet tai hallinneet XLIFF:ä, kääntämisen ammattilaiset käyttivät vähän XLIFF:ä sekä XLIFF:n suodattimien vähyys. Jatkokysymyksenä tutkimuksessa selvitettiin että mitä he olisivat halukkaita muuttamaan XLIFF:ssä. Vastauksena saatiin merkkeämisen yksinkertaistamista. Tarkempaa vaatimusten noudattamista. Tärkeän metadatan selventämistä. Parempaa työnkulun hallintaa. Tukea monikieliselle sisällölle. Työkaluja, jotka käyttävät ja luovat dataa eikä pelkästään tuo dataa XLIFF-formaatista. Kyselyn lopuksi vastaajilta kysyttiin heidän yleistä mielipidettä XLIFF-standardista. "Erittäin hyvänä" standardia piti 24 % vastanneista. 59 % vastaajista taas koki standardin "hyväksi". loput 17 % taas piti standardia "keskivertona". Huomion arvoista on että kukaan vastanneista ei valinnut vaihtoehdoksi "ei hyvä" vaan kaikki vastanneista piti standardia vähintäänkin keskivertoisena. (Anastasiou 2010)

4 YHTEENVETO

Tässä tutkielmassa käsiteltiin sekä sisällön lokalisointia että siihen läheisesti liittyviä standardeja. Sisällön lokalisointia käsitteenä kuvattiin siten että se on luonnollisen kielen kääntämistä kieleltä toiselle. Tällaisia sisältöjä voi olla esimerkiksi tekstidokumentit, pelien käyttöliittymät tai ohjekirjat. Useimmiten sisällön lokalisoimisessa käytetään tietokoneohjelmaa (kuten esim. Swordfish) käännöstyön apuna. Huomioitava on että ihmistä tarvitaan käännöstyössä edelleen, sillä tietokone ei pysty täydellisesti kääntämään sisältöä kohdekielelle. Sisällön lokalisoimisen standardit sekä tietokoneohjelmat tarjoavat kuitenkin huomattavan avun käännöstyön tekemiseen. Sisällön lokalisoimisen kenttää avattiin sisällön hallinnan mallilla, joka kuvaa sisällön lokalisoimisen eri komponentit eli osa-alueet. Sisällön lokalisoimisen keskeisiä käsitteitä kuten standardit, tekniikat ja käännösmuisti sekä syyt siihen miksi sisältöä lokalisoidaan on esitelty. Sisällön lokalisoimisen standardeista on mainittuna XLIFF-standardin lisäksi: TMX, TBX, URX, SRX, OLIF, xml:tm sekä GMX. Tässä eivät kuitenkaan ole kaikki sisällön lokalisoimiseen tarkoitetut standardit vaan ne on valittu vertilaukohtaksi XLIFF-standardille sekä antamaan yleiskuva siitä millaisia standardeja sisällön lokalisoimiseen on tarjolla. Pääpaino tämän tutkielman standardeista oli kuitenkin XLIFF-standardilla. XLIFF-standardin syntyä ja sen olemassaolon syitä on esitetty sekä tapoja sen käyttämiseen on havainnollistettu ja esitelty. Lisäksi Anastasioun (2010) tutkimuksen tulokset on esitelty tässä tutkielmassa, jotta lukijalle saadaan kuva siitä miltä XLIFF-standardin käyttö alan toimijoista tuntuu.

Sisältöjen localisointi on hyödyllistä, kun sisältöä tuottava organisaatio haluaa toimia muuallakin kuin kotimaassaan. Globalisoituvassa maailmassa tämä ei ole harvinaista, joten organisaatioiden joilla on toimintaa myös ulkomailla, tulisi harkita sisällön localisointia osana globalisoitumisensa strategiaa. Näin ollen kysymykseen: "Miksi sisältöjä localisoidaan?" voisi todeta että hyvän käyttäjäkokemuksen takia ja menestyksekkään kansainvälistymisen mahdollistamiseksi.

Reynoldsin ja Jewtushenkon (2005) mukaan XLIFF-standardi on nouseva standardi, joka tarjoaa tavan vaihtaa tietoa datan localisoimiseksi. XLIFF on

monien suurten organisaatioiden käytössä ja siksi sitä kehitetään aktiivisesti OASIS-järjestön toimesta. Edellä mainitusta syystä XLIFF-standardin ympärille luodaan koko ajan lisää työkaluja, joiden avulla sisällön lokalisoiminen tulisi helpottua entisestään.

Foster (2004) toteaa että yksi syy XLIFF-standardin käyttöön SUN Microsystemsillä on se, että sen avulla käännösmuisti- ja konekäännös työkalut voivat tehdä ehdotuksia kääntäjälle siitä, miten käsiteltävää dataa voidaan mahdollisesti kääntää. Tässä kohtaa huomataan, että pieni ominaisuus standardissa voi tehdä siitä kilpailukyvyltään erinomaisen tai suositumman mikäli ominaisuus on käyttäjän mielestä erinomainen ja siitä on hyötyä joko käyttäjälle tai organisaatiolle.

Tutkielman yhtenä tarkoituksena oli selvittää onko XLIFF-standardi hyvä sisällön lokalisoimisen standardi ja vastauksena voisi todeta, että XLIFF-standardi voi olla hyvä valinta organisaation sisällön lokalisointiin. Kuitenkin on huomioitavaa, että yrityksen tarkoituksena on otettava huomioon sopivaa standardia sovellettaessa, sillä jokaisella standardilla on ominaispiirteitä, jotka vaikuttavat standardin soveltuvuuteen tarkoitukseensa. Tutkimuskysymykseen: "Mitkä ovat XLIFF-standardin ominaispiirteet?" voisi todeta että, XLIFF-standardi kykenee hallitsemaan useita erilaisia tiedostomuotoja käännöstyössä sekä se antaa kääntäjälle vapauden keskittyä ainoastaan käännöstyöhön.

XLIFF-standardi on tiedetty toimivaksi hyvin sisällön lokalisoinnissa. Se ei ole yliveräinen standardi verrattuna muihin sisällön lokalisoimisen standardeihin, mutta käyttäjiä ja suosiota saadessaan siitä voi helposti kehittyä sisällön lokalisoinnissa merkittävä standardi mikäli se ei sitä jo ole. Standardoimista kuitenkin hankaloittaa monet muut käytettävissä olevat standardit ja siksi yleisen edun mukaisesti voisi kuvitella olevan järkevintä kehittää yhdessä standardi, jolla pyrittäisiin vastaamaan kaikkien sisältöä lokalisoivien tahojen tarpeisiin. Toisaalta tämä ajatus lienee ollut useimman standardin takana kunnes on huomattu että jokin ominaisuus puuttuu mikä taas on johtanut uuden standardin kehittämiseen. TMX-standardi lienee suosituin sisällön lokalisoimisen standardi, jota käytetään maailmalla yleisesti. Koska XLIFF pohjautuu monelta osin TMX-standardiin, sillä voi hyvinkin olla edellytykset syrjäyttää TMX-standardi suosituimpana. Kuten aikaisemmassa luvussa mainittu tutkimus toteaaakin niin työtä kuitenkin on tehtävänä XLIFF-standardissa, jotta se vastaisi kaikkien käyttäjien tarpeisiin. Lisäksi koulutusta olisi lisättävä siten että mahdollisimman moni sisällön lokalisoimisen parissa työskentelevä hallitsisi standardin ja näin ollen pystyisi työskentelemään sen avulla. Tämä myös lisäisi työkalujen kehittämistä siten että ne tukisivat XLIFF-standardia laaja-alaisemmin. Tavoitetilana voisikin pitää sitä mitä Anastasioun (2010) tutkimuksessa todetaan eli työkaluja, jotka luovat ja käsittelevät dataa XLIFF-standardin mukaisesti. XLIFF-standardista on kehitteillä versio 2.0, joka todennäköisesti tulee vastaamaan suurimpaan osaan sen aiemmin havaittuihin puutteisiin. Näin ollen jatkotutkimusaiheena voisi tulevaisuudessa olla XLIFF

2.0-standardin vaikutuksia XLIFF-standardin käyttöön ja sisällön lokalisoinnin kehitykseen.

LÄHTEET

- Anastasiou, D (2010). Survey on the Use of XLIFF in Localisation Industry and Academia. Haettu 30.10.2012 osoitteesta
<<http://www.cngl.ie/drupal/sites/default/files/papers2/Survey%20on%20the%20Use%20of%20XLIFF%20in%20Localisation%20Industry%20and%20Academia.pdf>>
- Bray, T. Paoli, J. Sperberg-McQueen, C. M. Maler, E. Yergeau, F. Cowan, J. (2006) Extensible Markup Language (XML). W3C.
- Bray, T. Hollander, D. Layman, A. (1998) Namespaces in XML World Wide Web Consortium Working Draft. 19.9.1998
- Dunham, R. (1991). Business design technology: "software development for customer satisfaction," *System Sciences, 1991. Proceedings of the Twenty-Fourth Annual Hawaii International Conference on*, vol.iii, no., pp.792-798 vol.3, 8-11 Jan 1991
- Esselink, B. (2000). A Practical Guide to Localization. John Benjamins Publishing.
- Foster, T. Using Translation Technology at Sun Microsystems.
Haettu 30.10.2012 osoitteesta:
<http://192.9.162.65/dev/gadc/technicalpublications/whitepapers/translation_technology_sun.pdf>
- Gommans, M. Krishnan, K. S. Scheffold, K. B. (2001) From Brand Loyalty to E-Loyalty: A Conceptual Framework. *Journal of Economic and Social Research* 3(1) 2001, 43-58
- Honkaranta, A. & Jauhiainen, E. 2006. Two Methods for Schema Design for Intelligent XML Documents in Organizations. *Technologies for Business Information Systems*. W. Abramowicz and H. C. Mayr. Dordrecht, Springer, 173-182.
- Hudik, T. Ruopp, A. (2011) The Integration of Moses into Localization Industry Haettu 30.10.2012 osoitteesta: <<http://mt-archive.info/EAMT-2011-Hudik.pdf>>
- Lewis, D. Curran, S. Feeney, K. Etzioni, Z. Keeney, J. Way, A. Schäler, R. Web Service Integration for Next Generation Localisation. *Workshop on Software Engineering, Testing, and Quality Assurance for Natural Language Processing (SETQA-NLP 2009) at NAACL-HLT 2009*, (s. 47 - 55) Boulder, Colorado <<http://mt-archive.info/EAMT-2011-Hudik.pdf>>
- Localization Industry Standards Association (LISA). (2011). Global information management Metrics eXchange (GMX) Haettu 30.12.2012 osoitteesta: <<http://web.archive.org/web/20110102011136/http://www.lisa.org/Global-information-m.104.0.html>>
- Localization Industry Standards Association (LISA). (2008). Segmentation Rules eXchange (SRX). Haettu 13.12.2012 osoitteesta:

- <<http://web.archive.org/web/20090220113311/http://www.lisa.org/Segmentation-Rules-e.40.0.html>>
- Localization Industry Standards Association (LISA). (2008). Systems to manage terminology, knowledge, and content - TermBase eXchange (TBX). Haettu 14.1.2013 osoitteesta: <http://www.galaglobal.org/oscarStandards/tbx/tbx_oscar.pdf>
- McCormick, S. Lieske, C. Culum, A. (2004). OLIF v.2: A Flexible Language Data Standard
- Miller, R. (2006). Multilingual Content Management: found in Translation. *E-content*. Haettu 30.10.2012 osoitteesta: <<http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=16786>>
- Mäki, S. (2008). Web 2.0 –verkkosovellusten piirteet ja Ajax niiden toteutuksessa. Tietojenkäsittelyopin pro gradu-tutkielma. Tampereen yliopisto
- OASIS. (2012). About us. Haettu 16.11.2012 osoitteesta <https://www.oasis-open.org/org>
- Okura, S. Bond, F. Yamamoto, Y. Murata, T. Uchimoto, K. Kato, M. Shimazu, M. Suzuki, T. (2009). "Introduction to UTX, a Specification for a Shared User Dictionary" IWIC '09 Proceedings of the 2009 international workshop on Intercultural collaboration (s. 147-154) ACM, New York, USA.
- Opentag.com. TMX. Haettu 12.1.2013 osoitteesta: <http://www.opentag.com/tmx.htm>
- Reynolds, P., Jewtushenko, T. (2005). What Is XLIFF and Why Should I Use It?. XML journal:
- Riikonen, M. (2006). Ohjelmistojen lokalisointi ja kansainvälistäminen. Tietojenkäsittelytieteen pro gradu -tutkielma. Jyväskylän yliopisto. <http://xml.sys-con.com/node/121957?page=0,3>
- Raya, R. (2007). XML in localisation: Reuse translations with TM and TMX. Haettu 30.10.2012 osoitteesta: <<http://www.ibm.com/developerworks/library/x-localis3/>>
- Raya, R. (2004). XML in localisation: Use XLIFF to translate documents. Haettu 30.10.2012 osoitteesta: <<http://www.ibm.com/developerworks/xml/library/x-localis2/>>
- Salminen, A. (2008). Towards Digital Government by XML Standardization: Methods and Experiences. *Proceedings of the XML Finland 2003* (s. 5-15)
- Salo, A. (2009). PHP- ja J2ME- sovelluksen lokalisointi. Tietojenkäsittelyn koulutusohjelma opinnäytetyö. Tampereen ammattikorkeakoulu.
- Savourel, Y. Schnabel, B. Jewtushenko, T. Domeny, D. (2006) XLIFF 1.2 Representation Guide for HTML. Haettu 5.1.2013 osoitteesta: http://docs.oasis-open.org/xliff/v1.2/xliff-profile-html/xliff-profile-html-1.2-cd02.html#elem_img
- Savourel, Y. Lommel, A. (2005) TMX 1.4b Specification. Haettu 10.1.2013 osoitteesta: <http://www.ttt.org/oscarStandards/tmx/>

Savourel, Y. Reid, J. Jewtushenko, T. Raya, R. M. (2008) XML Localisation Interchange File Format (XLIFF) 1.2 Specification Haettu 30.10.2012 osoitteesta:

<<http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html>>

Swordfish III User Guide. (2007-2012) Haettu 12.1.2013 osoitteesta:
<http://www.maxprograms.com/support/swordfish.pdf>

Zydroń, A. Raya, R. Bogacki, B. (2007). XML Text Memory (xml:tm) 1.0 Specification