



**TURUN
YLIOPISTO**
Kauppakorkeakoulu

Luonnollisen kielen käsittelyn hyödyntäminen toimitusketjujen kartoittamisessa

Toimitusketjujen johtamisen kandidaatintutkielma

Laatija:

Lasse Nordqvist

Ohjaaja:

KTT Sini Laari

17.4.2024

Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

Kandidaatintutkielma

Oppiaine: Toimitusketjujen johtaminen

Tekijä: Lasse Nordqvist

Otsikko: Luonnollisen kielen käsittelyn hyödyntäminen toimitusketjujen kartoittamisessa

Ohjaaja: KTT Sini Laari

Sivumäärä: 52 sivua

Päivämäärä: 17.4.2024

Toimitusketjujen läpinäkyvyyden merkitys on kasvanut viime vuosina merkittävästi, kun yritysten toimitusketjuista tulee yhä kansainvälisempiä, monimutkaisempia ja monikerroksisimpia. Yritysten arvonluonti toteutuu kasvavissa määrin toimitusketjun ylävirrassa, jolloin yritykset ovat yhä riippuvaisempia toimittajiensa suorituskyvystä. Vastuullisuuden merkityksen kasvu myös asettaa paineita toimitusketjun läpinäkyvyyden kasvattamiselle. Toimitusketjujen kasvaessa niiden luomaa verkostoa on vaikea hallita, jolloin yrityksen kyky reagoida toimitusketjussa tapahtuviin häiriöihin heikentyy merkittävästi.

Toimitusketjujen kartoittaminen toimii strategisena työkaluna toimitusketjun luoman verkoston tarkasteluun ja sen parempaan ymmärtämiseen. Kartoittamisen avulla voidaan luoda arvokasta liiketoimintatietoa, jonka pohjalta voidaan tehdä parempia strategisia päätöksiä toimitusketjuun liittyen ja parantaa toimitusketjun resilienssiä. Toimitusketjujen kartoittaminen on kuitenkin selkeästä hyödystään huolimatta vähäisessä käytössä toimitusketjujen johtamisessa. Tämä johtuu esimerkiksi kartoittamista varten kehitettyjen työkalujen puutteesta. Toimitusketjujen monimutkainen luonne tekee myös kartoittamisprosessista monimutkaisen, minkä takia toimitusketjukartan luominen vaatii organisaatiolta merkittävää työpanosta. Luonnollisen kielen käsittelyyn perustuva teknologia pyrkii vastaamaan työkalujen puutteeseen ja vähentämään kartoittamiseen vaadittavaa työpanosta merkittävästi.

Tämän tutkielman tavoitteena on tuoda esille yleisesti toimitusketjujen kartoittamisen hyödyntämistä organisaatioissa, keskittyen kuitenkin erityisesti luonnollisen kielen käsittelyn hyödyntämiseen kartoittamisen toteuttamisessa. Tutkimuskysymysten avulla pyritään tuomaan esille teknologian käyttöönottoon tulevaisuudessa liittyviä vaatimuksia organisaatioille ja tuodaan esille automaattisen kartoittamisen hyötyjä verrattuna perinteiseen kartoittamiseen.

Tutkimuskysymykset:

1. Mitkä luonnollisen kielen käsittelyn menetelmät soveltuvat hyödynnettäväksi toimitusketjujen kartoittamisessa?
2. Mitä edellytyksiä yrityksiltä vaaditaan toimitusketjujen kartoittamisen automatisoimiseksi luonnollisen kielen käsittelyn avulla?
3. Mikä on automatisoidun kartoittamisen hyöty verrattuna perinteiseen kartoittamiseen?

Julkisen datan määrän kasvaessa ja käytettävän teknologian kehittyessä, edellytykset toimitusketjujen kartoittamisen automatisoinnille luonnollisen kielen käsittelyn avulla parantuvat. Parantuneet edellytykset lisäävät saatuja hyötyjä ja käyttökohteita. Tällä hetkellä kehitetyt menetelmät eivät kuitenkaan pysty vielä luomaan monimutkaisia toimitusketjukarttoja. Teoria kehityksen pohjalle on kuitenkin luotu ja tulevaisuudessa teknologian kehittyessä tekstilähteiden pohjalta voidaan luoda monimutkaisiakin toimitusketjukarttoja.

Avainsanat: toimitusketjujen kartoittaminen, toimitusketjujen läpinäkyvyys, toimitusketjujen resilienssi, luonnollisen kielen käsittely, tiedonpöiminta

SISÄLLYS

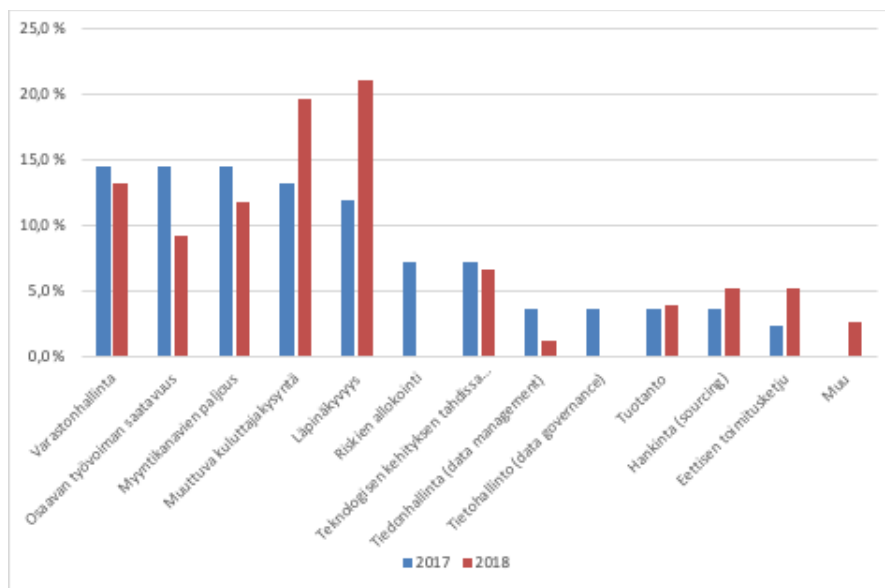
1	Johdanto	7
2	Toimitusketjujen kartoittaminen työkaluna toimitusketjujen johtamisessa	9
2.1	Toimitusketjujen kartoittaminen käsitteenä	9
2.2	Toimitusketjujen kartoittamisen hyödyt	9
2.3	Haasteet toimitusketjujen kartoittamisessa	11
2.3.1	Datan liiallinen yksityiskohtaisuus ja sen vaikutus vallitsevaan tilanteeseen toimitusketjussa	11
2.3.2	Liian yksityiskohtaisen kokonaiskuvan luonti	12
2.3.3	Kartan mittakaavan valitseminen	14
2.4	Toimitusketjukartan sisältö ja kartoittamisprosessi	14
2.5	Kartoittaminen eri toimitusketjun osissa	17
3	Luonnollisen kielen käsittely	20
3.1	Luonnollisen kielen käsittelyn määritelmä ja taustaa	20
3.2	Luonnollisen kielen käsittelyn prosessi	22
3.3	Luonnollisen kielen käsittelyn jälkeinen tiedon visualisointi	24
4	Luonnollisen kielen käsittelyn hyödyntäminen toimitusketjujen kartoittamisessa	26
4.1	Toimitusketjujen kartoittamiseen soveltuvat luonnollisen kielen käsittelyn menetelmät	26
4.2	Haasteet luonnollisen kielen käsittelyyn perustuvassa toimitusketjujen kartoittamisessa	31
4.2.1	Toimitusketjujen monimutkaisesta luonteesta aiheutuvat haasteet	31
4.2.2	Rajallisesta tiedon saatavuudesta aiheutuvat haasteet	34
4.3	Vaatimukset yrityksille toimitusketjujen kartoittamisen automatisoimiseksi luonnollisen käsittelyn avulla	35
4.4	Automaattisen kartoittamisen hyödyt verrattuna perinteiseen kartoittamiseen	38
5	Yhteenveto ja johtopäätökset	40
	Lähteet	43

KUVIOT

Kuvio 1: Suurimmat toimitusketjujen haasteet maailmanlaajuisesti vuosina 2017–2018 (Statista 2018)	7
Kuvio 2: Liian monimutkainen toimitusketjukartta (Farris 2010, 172)	13
Kuvio 3: Yksikertainen kartta H&M:n toimitusketjuverkostosta (McCarthy ym. 2022, 13)	14
Kuvio 4: Tekstiiliyrityksen keskivirran arvovirtakuvaus (Kumar 2016, 27)	18
Kuvio 5: Green Wolf -panimon alavirran jakelukartta (Sourcemap 2020)	19
Kuvio 6: Luonnollisen kielen käsittelyn prosessi (Dale 2010, 4; Li ym. 2021, 570; Sharma ym. 2022, 259)	23
Kuvio 7: Ontologiakartan hahmotelma (Liu ym. 2013, 118)	25
Kuvio 8: Esimerkki tekstistä löydetyistä ostaja-toimittajasuhteesta	27
Kuvio 9: Takaisinkytketyn neuroverkon ja muuntajan toiminnan hahmotelma	31
Kuvio 10: Transitiivisuuden ongelman hahmotelma (Wichmann ym. 2018, 1729, Wichmann 2022, 132)	33
Kuvio 11: Esimerkki yksittäisen tekstilähteen annotoinnista (Wichmann ym. 2020, 5329, Wichmann 2022, 111)	37
Kuvio 12: Tekstiperustaisen luonnollisen kielen käsittelyn markkinan koon muutos vuosina 2020–2030 (Statista 2023)	42

1 Johdanto

Kansainväliset toimitusketjujohtajat mainitsevat vuosina 2017–2018 toteutetussa tutkimuksessa toimitusketjun läpinäkyvyyden olevan yleisin haaste, jonka he kohtaavat toimitusketjujen johtamisessa. Vuonna 2018 21,1 % vastaajista oli kohdannut toimitusketjuissaan haasteita sen läpinäkyvydessä (Kuvio 1.) Viime vuosina koronapandemian aiheuttamat toimitusketjujen haasteet ovat kasvattaneet näkyvyyden (Mubarik ym. 2023, 2653) ja muiden toimitusketjujen resilienssiä parantavien toimien merkitystä (Ivanov & Dolgui 2020, 2904). Toimitusketjujen kartoittaminen parantaa niiden läpinäkyvyyttä ja resilienssiä (Mubarik ym. 2021, 7).



Kuvio 1: Suurimmat toimitusketjujen haasteet maailmanlaajuisesti vuosina 2017–2018 (Statista 2018)

Ensisijaisesti kartoittaminen toimii työkaluna toimitusketjuun liittyvän informaation kommunikaatioon ja selkeyttämiseen (Gardner & Cooper 2003, 37, 39). Selkeän informaation pohjalta yrityksissä voidaan tehdä paremmin toimitusketjuihin liittyviä strategisia päätöksiä (Gardner & Cooper 2003, 39–40) ja paremmin välttää mahdolliset toimitusketjun häiriöt (Mubarik ym. 2023, 2653). Toimitusketjujen kartoittamisen akateemista tutkimusta on kuitenkin sen tärkeyteen verraten melko vähän. Erityisesti toimitusketjujen kartoittamisen konkreettisen lisäarvon mittaamiseen pyrkivässä tutkimuksessa on puutetta. (Fabbe-Costes ym. 2020, 1475–1476; Mubarik ym. 2023, 2654.) Vaikka toimitusketjujen kartoittamisen on todistettu olevan toimiva keino niiden strategisessa johtamisessa, kartoittamisen käyttö liiketoiminnassa on vielä vähäistä.

Toimitusketjujen kartoittamiseen tarkoitettuja työkaluja tulee kehittää, jotta tulevaisuudessa sen laajempi käyttöönotto liiketoiminnassa on mahdollista (Fabbe-Costes ym. 2020, 1486; Farris & Ahmed 2019, 156–157.)

Luonnollisen kielen käsittelyn vaatiman luonnollisen kielen muodossa olevan datan määrän kasvu (Grosman ym. 2020, 1; Li ym. 2021, 569; Wichmann ym. 2018, 1726) ja teknologiset edistykset syväoppimisessa ja luonnollisen kielen käsittelyssä (Wichmann ym. 2020, 5321) kannustavat yrityksiä kehittämään luonnollisen kielen käsittelyyn perustuvia ratkaisuja toimitusketjujen kartoittamisen automatisoimiseksi. Luonnollisen kielen käsittelyn taustalla ovatkin nykyään vahvasti syväoppimisen periaatteet ja metodit (Deng & Liu 2018, 6–10). Syväoppimisen käytöstä toimitusketjujen kartoittamisessa on akateemista tutkimusta kuitenkin vähän verrattuna muihin toimitusketjujen johtamisen aihealueisiin. Hosseinia Shavaki & Ebrahimi Ghahnavieh (2022) syväoppimisen käyttöä toimitusketjujen johtamisessa tarkastelevassa systemaattisessa kirjallisuuskatsauksessa todetaan ainoastaan 2 % aihepiirin akateemisesta tutkimuksesta käsittelevän toimitusketjujen kartoittamista.

Toimitusketjujen kartoittamisen aihealuetta tulee siis tutkia ja kehittää, jotta sen laajamittaisempi käyttöönotto liiketoiminnassa on mahdollista tulevaisuudessa. Luonnollisen kielen käsittelyn teknologialla on havaittu olevan potentiaalia tämän kehityksen edistämiseen. Tämän tutkielman tarkoituksena on siis tarkastella toimitusketjujen kartoittamista ja luonnollisen kielen käsittelyn hyödyntämistä sen automatisoinnissa. Tutkimuskysymyksillä pyritään tuomaan esille, mitä organisaatioilta vaaditaan, jotta se voisi tulevaisuudessa hyödyntää luonnollisen kielen käsittelyä toimitusketjujen kartoittamisessa ja automaattisen kartoittamisen hyödyistä verrattuna perinteiseen kartoittamiseen.

Tutkimuskysymykset:

1. Mitkä luonnollisen kielen käsittelyn menetelmät soveltuvat hyödynnettäväksi toimitusketjujen kartoittamisessa?
2. Mitä edellytyksiä yrityksiltä vaaditaan toimitusketjujen kartoittamisen automatisoimiseksi luonnollisen kielen käsittelyn avulla?
3. Mikä on automatisoidun kartoittamisen hyöty verrattuna perinteiseen kartoittamiseen?

2 Toimitusketjujen kartoittaminen työkaluna toimitusketjujen johtamisessa

2.1 Toimitusketjujen kartoittaminen käsitteenä

Toimitusketjujen kartoittamisella tarkoitetaan yrityksen toimitusketjun muodostaman verkoston visualisointia toimitusketjuun liittyvän informaation selkeyttämiseksi. Kartat voivat olla luonteeltaan deskriptiivisiä ja kuvata toimitusketjun nykytilannetta tai preskriptiivisiä ja kuvata toimitusketjun mahdollista koostumusta tulevaisuudessa. Toimitusketjukartat luovat pohjaa oman organisaation toimitusketjun ymmärtämiseen ja niiden avulla toimitusketjuun liittyviä strategisia päätöksiä voidaan kohdistaa paremmin. Käytännössä kartoittamisen tavoitteena on luoda toimitusketjukartta, joka laajentaa yritysten kykyä havainnoida toimitusketjun eri osissa tapahtuvaa toimintaa. (Barroso ym. 2011, 169–173; Gardner & Cooper 2003, 37, 39.)

Toimitusketjujen kartoittamista ei tule sekoittaa arvovirtakuvaukseen (eng. value stream mapping) tai prosessien kartoittamiseen, jotka ovat operatiivisen ja taktisen tason liikkeenjohdon työkaluja mutta myös niillä on omat käyttökohteensa toimitusketjujen kartoittamisessa. (Gardner & Cooper 2003, 45–46; MacCarthy ym. 2022, 3.) Arvovirtakuvaus pyrkii keskittymään ainoastaan toimintaan, joka lisää arvoa lopputuotteeseen, kun taas toimitusketjun kartoittamisessa pyritään kuvaamaan yleisellä tasolla kaikkia toimijoiden toimitusketjuun liittyvää toimintaa (Hines & Rich 1997, 46; MacCarthy ym. 2022, 15). Prosessien kartoittamisessa taas keskitytään yleisesti yrityksen sisäisiin prosesseihin. Prosessien kartoittamista ja arvovirtakuvausta voidaankin käyttää hyväksi toimitusketjun keskivirran kartoittamiseen. (Gardner & Cooper 2003, 45; Klotz ym. 2008, 624; Mubarik. ym 2021, 2.)

2.2 Toimitusketjujen kartoittamisen hyödyt

Yksi merkittävistä toimitusketjun kartoittamisen eduista on sen kyky edistää toimitusketjujen läpinäkyvyyttä ja resilienssiä (Mubarik ym. 2021, 7; Norwood & Peel 2021, 417–418; Wichmann ym. 2022, 76–77). Toimitusketjujen resilienssillä viitataan organisaation kykyyn ylläpitää toimitusketjunsuorituskykyä mahdollisten häiriöiden kohdistuessa siihen ja sen kykyyn palauttaa sen toiminta normaaliksi (Christopher & Peck 2004, 4; Ponomarov & Holcomb 2009, 131). Toimitusketjujen läpinäkyvyys taas viittaa siihen, kuinka hyvin toimitusketjun yksittäisillä toimijoilla on saatavillaan

sellaista toimitusketjuun liittyvää tietoa, jota ne pitävät olennaisena oman toiminnan kannalta ja kuinka hyvin ne jakavat toimitusketjuun liittyvää tietoa, josta he uskovat toimitusketjun muiden toimijoiden hyötyvän (Barratt & Oke 2007, 1230).

Toimitusketjujen kartoittaminen ja sen tuoma läpinäkyvyys ovat olennaisia toimitusketjun vastuullisuuden kannalta (Al-Khatib 2023, 1620–1621; Mubarik ym. 2021, 2; Norwood & Peel 2021, 424; Wichmann ym. 2022, 77). Toimitusketjun vastuullisuutta voidaan käyttää monella tavalla taloudellisen arvon luomiseksi organisaatiolle (Mefford 2011, 135).

Kartoittaminen myös helpottaa uusien toimijoiden integrointia toimitusketjuun. Kun yrityksen toimitusketjusta on olemassa visualisointi toimitusketjukartan avulla, uudet toimijat pystyvät paremmin sisäistämään oman roolinsa toimitusketjussa ja suhteuttamaan toimintaansa tähän. Kartta auttaa myös nykyisten toimijoiden toiminnan kehittämisessä. (Fabbe-Costes ym. 2020, 1487; Gardner & Cooper 2003, 40.)

Toimitusketjukartta helpottaa toimitusketjun ymmärryksen lisäämistä myös oman organisaation sisällä koko organisaation laajuisesti ja auttaa tunnistamaan kehityskohteita oman toimitusketjun toiminnasta (Barroso ym. 2011, 172–173; Fabbe-Costes ym. 2020, 1487–1488; Farris & Ahmed 2019, 148; Norwood & Pell 2021, 416–417).

Toimitusketjukartta myös edistää eri osastoiden toimitusketjuun liittyvien tavoitteiden yhteensovittamista (Fabbe-Costes ym. 2020, 1488). Toimitusketjusta vastaavissa organisaation toiminnoissa kartoittamisen hyöty liittyy sen kykyyn yhdistää toimitusketjujen aineelliset ja aineettomat tekijät yhteen helpommin ymmärrettävissä olevaan kokonaisuuteen (Fabbe-Costes ym. 2020, 1487). Toimitusketjuihin liittyy fyysisen tavaran liikkumisen ohjaamista mutta merkittävältä osin toimitusketjujen johtamiseen liittyy myös aineettomien ja abstraktien kokonaisuuksien hallintaa (Fabbe-Costes ym. 2020, 1487; Subramanian ym. 2015). Toimitusketjukartta voi auttaa myös liiketoimintatiedon hankinnassa, kun kartoittamisprosessin avulla saatetaan tunnistaa myös kilpailijoiden toimittajia ja asiakkaita (Wichmann 2022, 87). Toimitusketjukartan avulla voidaan myös tunnistaa toimittajien keskittymistä (Wichmann 2022, 87), mikä on merkittävä toimitusketjuriskin aiheuttaja (Inomata & Hanaka 2024, 177).

Suorien hyötyjen lisäksi kartoittamisella on myös epäsuoria hyötyjä.

Kartoittamisprosessi itsessään laajentaa siihen osallistuvien henkilöiden tietoisuutta ja

herättää keskustelua toimitusketjusta ja sen luomasta kokonaisuudesta.

Toimitusketjuihin liittyvän tietoisuuden kasvaessa strateginen intuitio ohjaa tekemään parempia toimitusketjuun liittyviä strategisia päätöksiä. (Fabbe-Costes ym. 2020; Gardner & Cooper 2003, 40; Norwood & Peel 2021, 424–426.) Strateginen intuitio on strategiatyöhön liittyvä aivoissa tapahtuva prosessi, jossa aiempia muistijälkiä aiheesta yhdistetään sopivan strategian luomiseksi. Strategisen intuition käyttö strategiatyössä parantaa organisaation suorituskykyä merkittävästi. (Bullini Orlandi & Pierce 2020, 135, 140; Songkajorn ym. 2022, 2, 13–14.)

2.3 Haasteet toimitusketjujen kartoittamisessa

Gardner & Cooper (2003) tunnistavat artikkelissaan neljä selkeää haastetta, joita organisaatiot saattavat kohdata toimitusketjujen kartoittamisessa: kartan sisältämän datan liiallinen yksityiskohtaisuus, kartan aiheuttamat tahattomat muutokset toimitusketjun vallitsevaan tilanteeseen, kartan sisällön liiallinen yksityiskohtaisuus ja organisaatiolle soveltuvan kartan mittakaavan valinta. Esimerkiksi Farris (2010) ja Farris & Ahmed (2019) ovat pyrkineet esittelemään ratkaisuja näihin haasteisiin. Toimitusketjujen monimutkainen ja samanaikaisesti fyysinen sekä abstrakti luonne (Fabbe-Costes ym. 2020, 1487; Subramanian ym. 2015) sekä kartoittamisessa hyödynnettävien työkalujen puute (Fabbe-Costes ym. 2020, 1486; Farris & Ahmed 2019, 156–157) tekevät kartoittamisesta monimutkaisen prosessin, joka vaatii suuren työpanoksen ja paljon aikaa.

2.3.1 Datan liiallinen yksityiskohtaisuus ja sen vaikutus vallitsevaan tilanteeseen toimitusketjussa

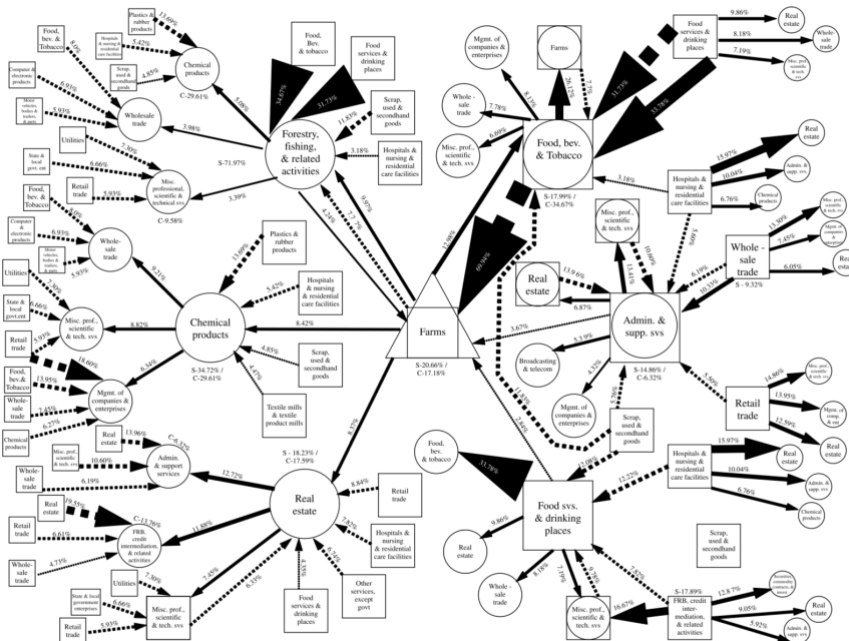
Liian yksityiskohtaista dataa sisältävä toimitusketjukartta saattaa paljastaa eri toimijoiden kilpailuetuja. Toimittajiin liittyvän datan hankinnassa ongelmaksi yleensä muotoutuu se, että toimitusketjun muut toimijat haluavat pitää toimittajiinsa liittyvän informaation yksityisenä. Tähän johtaa esimerkiksi se, että toimijoilta puuttuu positiivinen kannustin toimittajatiedon jakamiseen. Kuitenkin jotta toimitusketju toimii, eri toimijat tarvitsevat toimitusketjuun liittyvää informaatiota. Julkista ja ostettavissa olevaa dataa toimittajasuhteista on kuitenkin enenevässä määrin saatavilla, mikä helpottaa toimittajadatan hankintaa. Kerätyn datan pohjalta itse kartoittamisprosessissa tulee tarkasti miettiä, mitä kartassa halutaan esittää ja onko kartta muiden toimijoiden

saatavilla vai tarkoitettu vain organisaation omaan käyttöön. Kartoittamisessa tulisi olla tietoinen siitä, mitä vaikutuksia toimitusketjukartan sisältämällä datalla saattaa olla toimitusketjun vallitsevaan tilanteeseen ja toimijoiden välisiin suhteisiin. Oikein käytettynä toimitusketjukartta saattaa nostaa kaikkien toimijoiden tuottoja mutta se saattaa myös vääristää toimijoiden välistä kilpailutilannetta ja muuttaa tahattomasti toimitusketjun vallitsevaa tilannetta. (Farris 2010, 166–167; Farris & Ahmed 2019, 141; Gardner & Cooper 2003, 58; Norwood & Peel 2021, 418; Wichmann 2022, 76–77.)

2.3.2 Liian yksityiskohtaisen kokonaiskuvan luonti

Kartoittamisprosessissa tulee myös välttää liian yksityiskohtaisen kokonaiskuvan luomista organisaation toimitusketjusta. Liian yksityiskohtaisen toimitusketjukartan luomisesta ja ylläpitämisestä tulee siitä saavutettuun strategiseen hyötyyn verrattuna liian haastavaa. Toimitusketjukartan toimijoiden määrä kasvaa hyvin nopeasti, jos siitä halutaan luoda todenmukainen. Toimitusketjujen kartoittamisessa tulee pitäytyä helposti visualisoitavissa olevaan toimijoiden määrään ja keskittyä kartoittamisessa enemmän esimerkiksi toimitusketjun kannalta strategisesti tärkeimpiin toimijoihin. (Cooper ym. 1997, 79–81; Farris 2010, 169–170; Farris & Ahmed 2019, 142; Gardner & Cooper 2003, 58–59; Lambert ym. 1998, 5.) Kun toimitusketjukartta on liian monimutkainen, sen sisältämää informaatiota on hyvin vaikea sisäistää (Kuvio 2). Kartan sisältämän tiedon yksityiskohtaisuutta tärkeämpi tekijä on sen luotettavuus (Fabbe-Costes ym.

2020, 1486). Liian yksinkertaisen kartan luomista tulee kuitenkin välttää (Fabbe-Costes ym. 2020, 1483).



Kuvio 2: Liian monimutkainen toimitusketjukartta (Farris 2010, 172)

Kartan päivitettävyyttä liittyy olennaisesti yksinkertaistettuun toimitusketjun kokonaiskuvaan. Toimitusketjukartta tulee olla helposti päivitettävissä, kun toimijoiden väliset suhteet muuttuvat toimitusketjussa. (Fabbe-Costes ym. 2020, 1486; Gardner & Cooper 2003, 57.) Strategiset toimittajasuhteet kestävät keskimäärin noin 15 vuodesta ylöspäin (Dubois ym. 2021, 697; Li ym. 2022, 303) mutta organisaatioilla yleensä on myös strategisesti vähemmän tärkeitä toimittajasuhteita, joiden kestot ovat lyhyempiä (Dubois ym. 2021, 697). Toimijoiden välisten suhteiden päättyessä ne tulee mahdollisimman nopeasti poistaa toimitusketjukartasta, koska ainoastaan olemassa olevilla suhteilla on toimitusketjukartan visualisoinnin kannalta merkitystä (Wichmann ym. 2018, 1728). Koska strategisesti vähemmän merkittävät toimittajasuhteet kestävät vähemmän aikaa, niiden hallintaan ei ole kannattavaa allokoida yhtä paljon resursseja kuin strategisesti tärkeämpien toimittajien hallintaan. Strategisesti vähemmän merkityksellisten toimijoiden kohdalla yrityksen on luotettava niiden kykyyn hallita toimitusketjussa itselleen merkittäviä toimijoita (Lambert ym. 1998, 8). Jos yrityksellä

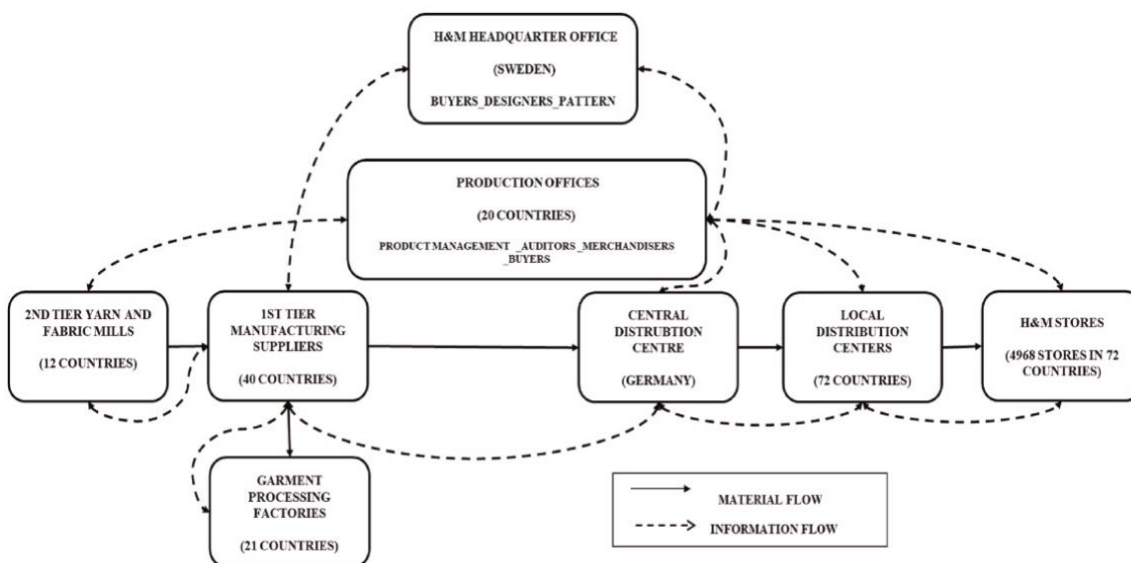
on kuitenkin jo suoria asiakkaita ja toimittajia hyvin paljon, on toimitusketjun syvempien tasojen hallinta ja kartoittaminen hyvin haastavaa (Lambert ym. 1998, 6).

2.3.3 Kartan mittakaavan valitseminen

Toimitusketjujen kartoittamisessa tulee myös arvioida, missä mittakaavassa toimitusketjukartta olisi kannattavaa toteuttaa, jotta se olisi organisaatiolle strategisesti kannattavin. Mittakaavat vaihtelevat globaalien arvoketjujen kartoittamisesta yritysten sisäisten prosessien kartoittamiseen. (MacCarthy ym. 2022, 5). Jos mittakaava on liian pieni, strategisesti merkittävää liiketoimintatietoa saattaa jäädä mittakaavan ulkopuolelle ja jos se on taas liian iso, strategisesti merkittävät yksityiskohdat saattavat jäädä huomaamatta. Jotta kartoittamisen strateginen painopiste säilyisi, on suositeltavaa aloittaa isommasta mittakaavasta ja kartoittamisen prosessin edetessä tarvittaessa siirtyä yksityiskohtaisempaan mittakaavaan. (Farris 2010, 177; Gardner & Cooper 2003, 59.)

2.4 Toimitusketjukartan sisältö ja kartoittamisprosessi

Kaikkien toimitusketjukarttojen tulisi sisältää vähintään tietoa toimitusketjuun osallistuvista toimijoista ja niiden välisistä yhteyksistä. Olennaisinta yhteyksistä on tietää, mitä kaikkea toimijoiden välisissä virroissa kulkee ja mihin suuntaan yhteyksien luomat virrat kulkevat. Toimijoiden välillä voi kulkea esimerkiksi materiaalia, tietoa tai rahaa. (Barroso ym. 2011, 171; Craighead ym. 2007, 134–135; Farris 2010, 171; MacCarthy ym. 2022, 1; Mubarik ym. 2021, 3; Wichmann ym. 2018, 1728.)



Kuvio 3: Yksikertainen kartta H&M:n toimitusketjuverkostosta (McCarthy ym. 2022, 13)

Näiden lisäksi toimitusketjukarttaan voidaan lisätä organisaation toimitusketjun toiminnan kannalta olennaisia ominaisuuksia. Toimitusketjukartta voi visualisoida toimijoiden maantieteellistä jakautumista (MacCarthy ym. 2022, 5; Mubarik ym. 2023, 2657; Wichmann ym. 2018, 1728, Wichmann 2022, 87). Toimijoita voidaan myös erotella sen mukaan, kuinka tärkeitä ne ovat organisaation toiminnalle, mikä niiden rooli on toimitusketjussa ja kuinka monta samankaltaista toimijaa toimitusketjussa on (Barroso ym. 2011, 171; Cooper ym. 1997, 79; Farris & Ahmed 2019, 144; Gardner & Cooper 2003, 40, 49–51; Lambert ym. 1998, 40). Jokaiselle toimijalle voidaan myös erikseen lisätä, mistä lähteestä siihen liittyvät tiedot on saatu ja kuinka uutta hyödynnetty tieto on (Barroso ym. 2011, 171, Wichmann 2022, 87). Materiaalivirtojen osalta voidaan myös mainita niistä vastaavat logistiikkatoimijat (Wichmann 2022, 77).

Kun toimijoista on kerätty tarvittavat tiedot, kartan avulla voidaan luoda visualisointi toimitusketjusta tietyn toimijan näkökulmasta. Yleisesti kartoittaminen tehdään oman organisaation näkökulmasta mutta kartoittamisprosessissa voidaan myös kartoittaa kilpailijoiden osallisuutta toimitusketjuun. Valitulla näkökulmalla on merkittävä vaikutus siihen, millaisia strategisia johtopäätöksiä toimitusketjukartan avulla voidaan tehdä. (Craighead ym. 2007, 134–135; Gardner & Cooper 2003, 40, 55; Lambert ym. 1998, 7–9.) Toimitusketjukartta voidaan tehdä myös jonkin tietyn organisaation valmistaman tuotteen näkökulmasta (Wichmann ym. 2018, 1728). Toimitusketjukartan olisi hyvä jollain tavalla huomioida muita toimijoita, jotka ovat näkökulmaksi valitun organisaation kanssa samalla tasolla eli käytännössä organisaation kilpailijoita. Kilpailijoilla on vaikutusta organisaation suhteisiin niin toimittajien kuin myös asiakkaiden kanssa. (Lambert ym. 1998, 8.)

Kartan tulisi olla sisällöltään universaalisti ymmärrettävissä, jotta tarvittaessa kaikki toimitusketjun toimijat pystyvät sisäistämään sen sisältämän informaation. Kartan sisältämän informaation tulkinta tulisi olla sama riippumatta siitä, missä tehtävässä tulkitsija toimii. Kartan tulee toimia eri osastojen välisenä toimitusketjuun liittyvän informaation kommunikaatiokeinona. Tämän takia toimitusketjukartan luomisprosessissa pääpaino on sen visuaalisessa ilmeessä. (Fabbe-Costes ym. 2020, 1479, 1486–1487; Gardner & Cooper 2003, 37, 40). Helposti ymmärrettävällä toimitusketjukartalla voidaan ohjeistaa toimitusketjun uusia toimijoita sen toiminnasta (Wichmann ym. 2022, 77).

Toimitusketjukartoista on kannattavaa luoda deskriptiivinen ja preskriptiivinen versio, koska toivottavaan tulevaisuuden tilaan vaadittuja toimia on helpompi havaita ja suunnitella, kun toimitusketjun nykytilaa voidaan verrata sen tulevaisuuden tilaan. Kahden kartan vertailu auttaa konkreettisen toimitusketjun muutossuunnitelman toteuttamisessa. Toimitusketjujen kartoittaminen on olennainen työkalu niin vähittäisen jatkuvan kehityksen kuin myös suurempien toimitusketjuihin liittyvien strategisten päätösten tukemisessa. (Gardner & Cooper 2003, 40.)

Gardner & Cooper (2003) mukaan toimitusketjujen kartoittamisesta saadaan paras strateginen hyöty silloin, kun kartoittamisprosessi on tarkkaan määritelty. Fabbe-Costes ym. (2020) toteuttama monimenetelmäinen case-tutkimus kuitenkin osoittaa, että toimitusketjujen kartoittaminen on hyvin subjektiivinen prosessi ja samasta toimitusketjusta tehtyjen karttojen sisältö saattaa vaihdella riippuen siitä, kuka sen on tehnyt. Näkemys ja suhde toimitusketjun kokonaisuuteen vaikuttavat siihen, miten toimitusketjun kartoittaa. Näkemyseroja saattaa syntyä esimerkiksi siinä, mihin asti organisaation kannalta olennaiset toimitusketjun osat ulottuvat. Tämä tekee systemaattisen kartoittamisprosessin määrittämisestä vaikeaa ja haastaa sen tarpeellisuutta. Eriävät näkemykset voivat tukea toisiaan ja toimitusketjukarttaa voitaisiin muokata sen mukaan, mikä sen käyttäjän suhde on toimitusketjuun.

Toimitusketjujen kartoittamisessa voidaan ottaa erityiseksi painopisteeksi toimitusketjun resilienssin parantaminen (Barroso ym. 2011, 171–173).

Toimitusketjukartan avulla toimitusketjusta voidaan löytää mahdollisia haavoittuvuuksia. Haavoittuvuuksien löytämisen jälkeen voidaan analysoida, mitä vaikutuksia niiden realisoitumisella olisi toimitusketjun toimintaan ja visualisoida realisoitumista kartoittamisen keinoin. (Barroso ym. 2011, 172–173; Norwood & Pell 2021, 416–417.) Jos toimitusketju ei ole resilientti kyseistä haavoittuvuutta kohtaan, tulisi sen vaikutusten lieventämiseksi määrittää erilaisia keinoja ja tarkastella kartoittamisen keinoin niiden vaikutusta toimitusketjuun (Barroso ym. 2011, 173). Erityisesti tietyllä alalla erikoistuneet toimitusketjut ovat hyvin haavoittuvia kyseiseen toimitusketjuun kohdistuville häiriöille. Kartoittaminen ja haavoittuvuuksien löytäminen pakottavat katsomaan oman alan toimitusketjun ulkopuolelle ratkaisujen löytämiseksi. Jos yksittäisen alan toimitusketjuissa on havaittavissa haavoittuvuuksia, ne voidaan paikata esimerkiksi yhdistämällä toisen alan toimitusketjun tarjoamia substituuotteja oman toimitusketjun tueksi. (Norwood & Peel 2021, 413–415, 421–424.)

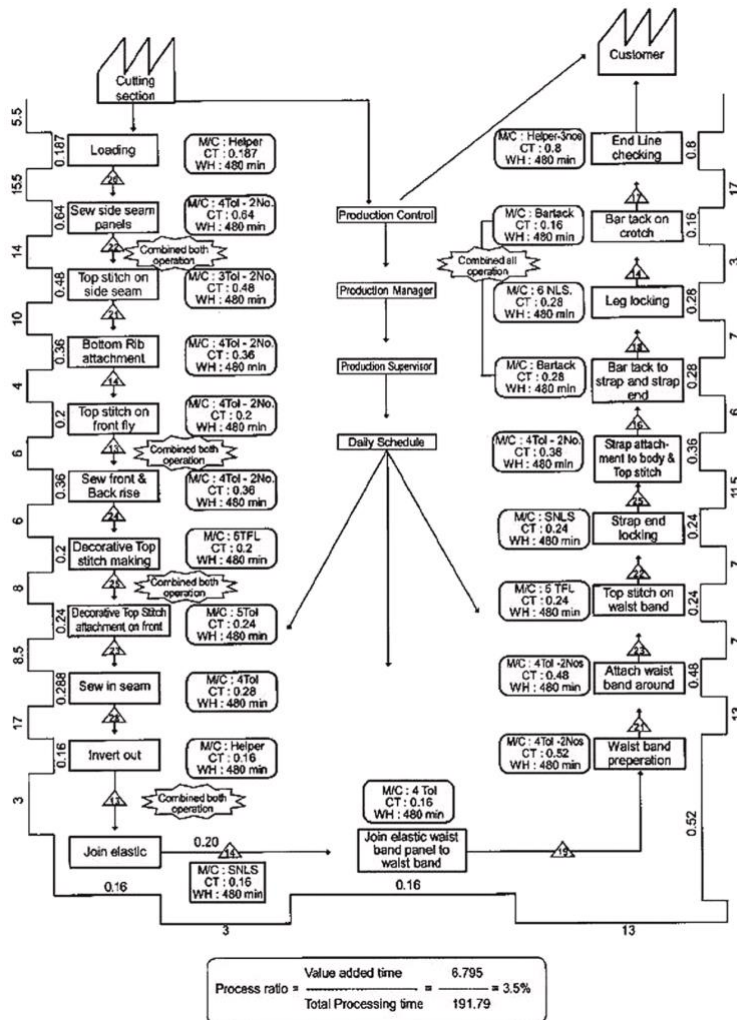
Kaikista erilaisista skenaarioista tulisi verrattavissa olevien mittareiden avulla, mitä vaikutuksia niillä on toimitusketjun suorituskykyyn. Esimerkiksi toimittajien kyky palautua haavoittuvuuden realisoitumisesta, organisaation kyky ylläpitää haluttua palvelutasoa, toimitusajat, haavoittuvuuden realisoitumisen ja lieventämiskeinojen aiheuttamat kustannukset soveltuvat suorituskyvyn ja kustannusten mittareiksi. (Barroso ym. 2011, 173; Gardner & Cooper 2003, 45; Norwood & Peel 2021, 418–421.)

2.5 Kartoittaminen eri toimitusketjun osissa

Toimitusketjun kartoittamisen avulla voidaan tarkastella toimintaa toimitusketjun ylä-, keski- ja alavirroissa. Toimitusketjun ylävirralla viitataan oman yrityksen toimittajaverkoston ja sen toimintaan, keskivirralla viitataan oman yrityksen omaan toimintaan ja alavirralla viitataan oman yrityksen asiakkaiden luoman verkoston toimintaan. (Kazemi & Szmerekovsky 2015, 111; Mubarik ym. 2021, 5; Mubarik ym. 2023, 2657–2658; Norwood & Peel 2021, 416.) Kartoittamisprosessissa voidaan keskittyä johonkin tiettyyn toimitusketjun osaan tai siinä voidaan pyrkiä luomaan kuvaa koko toimitusketjun pituudelta. Ylä- ja keskivirran kartoittaminen edistävät toimitusketjun resilienssiä mutta alavirran kartoittaminen ei ole yhtä olennaista resilienssin ylläpidossa. Kaikkien osien kartoittaminen kuitenkin edistää toimitusketjun läpinäkyvyyttä. (Mubarik ym. 2021, 7.) Eri toimitusketjun osissa kartoittamisessa tulee painottaa erilaisia tekijöitä. Mubarik ym. (2023) toteuttama aiheeseen liittyvä kyselytutkimuksen eksploratiivinen faktorianalyysi tarkastelee huomioonotettavia tekijöitä eri toimitusketjun osien kartoittamisessa.

Ylävirran kartoittamisen kannalta merkittävä tekijä ovat selkeästi määritellyt prosessit toimittajien kanssa toimimiseen, jotka helpottavat ylävirran suhteiden visualisointia (Gardner & Cooper 2003, 40). Selkeät suhteet toimittajien kanssa helpottavat reaaliaikaisen informaation saamista toimijoiden välisistä virroista. Eri toimijoiden maantieteellinen jakautuminen tulisi myös olla selkeästi esillä ylävirran toimitusketjukartassa. Vähintään oman toiminnan kannalta kriittisten komponenttien osalta tulisi myös tuntea toimittajat suorista toimittajista ylöspäin, joiden tuntemisen merkitys korostuu erityisesti toimitusketjuhäiriöihin varautumisessa (Lambert ym. 1998, 9). Tässä organisaatioilla on kuitenkin erityisesti parannettavaa (The Business Continuity Institute 2023, 55). Toisen tason toimittajien tunteminen on haastavaa erityisesti yrityksille, joilla on paljon suoria toimittajia (Lambert ym. 1998, 6).

Ylävirran toimitusketjukartan tulisi myös olla tarpeeksi pääpiirteinen ja yksinkertainen, jottei sen hallinnasta tulisi liian haastavaa (Barroso ym. 2011, 169; Cooper ym. 1997 79–81; Gardner & Cooper 2003, 58–59; Norwood & Peel 2021, 414).

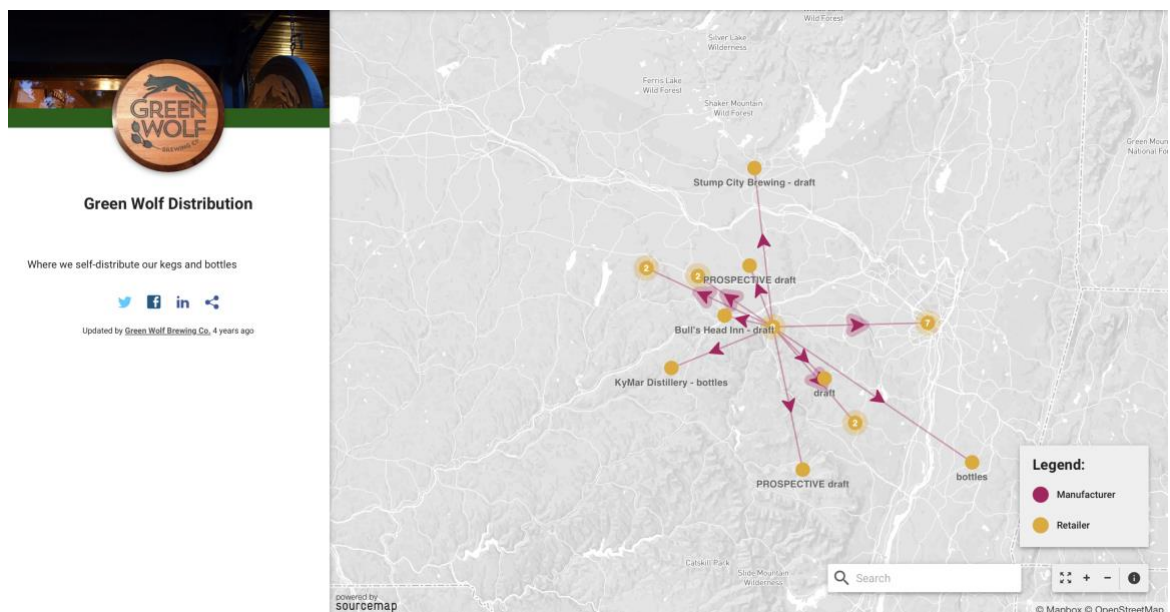


Kuvio 4: Tekstiiliyrityksen keskivirran arvovirtakuvaus (Kumar 2016, 27)

Keskivirran kartoittamisessa merkittävä tekijä on se, että yrityksen sisällä prosesseista vastaavat henkilöt saavat kartoittamisen avulla varoituksia, jos prosesseihin on syntymässä mahdollisia pullonkauloja. Keskivirrassa reaaliaikaista informaatiota on tärkeä saada prosessien tehokkuudesta ja virroista koko organisaatiossa leveydeltä, jotta mahdollisia tehottomuuden aiheuttajia voidaan korjata mahdollisimman nopeasti. Informaatio tulee myös olla nähtävissä kaikkialla organisaatiossa, jotta välttytään mahdollisilta informaatioasiiloilta (Gardner & Cooper 2003, 39–40; Hussain ym. 2017, 168; Lambert ym. 1998, 9; Weerakkody ym. 2011, 325). Informaatioasiilojen välttäminen on tärkeää, koska sama toimittaja saattaa osallistua organisaation sisällä monien eri yksiköiden prosesseihin (Lambert ym. 1998, 12). Prosessien kartoittamisessa

ja arvovirtakuvauksessa on kehitetty työkaluja reaaliaikaisen tilannekuvan ylläpitämiseen, mutta ne ovat rajoittuneet yksinkertaisiin prosesseihin ja virtoihin, eivätkä ne pysty reaaliaikaisesti ylläpitämään tilannekuvaa monimutkaisista virroista tai prosesseista (Benzi ym. 2018, 195; Huang ym. 2019, 2). Prosessien kartoittamisen ja arvovirtakuvauksen avulla organisaatiot pystyvät myös paremmin seuraamaan toimitusketjustrategiansa toteutumista operationaalisella ja taktisella liiketoiminnan tasolla. Niiden käyttö auttaa myös sopeutumaan dynaamisen liiketoimintaympäristön muutoksiin (Gardner & Cooper 2003, 39; Ivanov 2020, 10–11; Mubarik ym. 2020, 2657).

Mubarik ym. (2023) lisäksi myös Fabbe-Costes ym. (2020) tarkastelee tekijöitä, jotka ovat alavirran kartoittamisessa tärkeitä. Alavirran kartoittamisessa toimijoiden maantieteellisen jakautumisen selkeä esittäminen on ylävirtaa merkittävämmässä roolissa (Fabbe-Costes ym. 2020, 1483; MacCarthy ym. 2022, 12–13; Mubarik ym. 2023, 2664–2665; Kuvio 5).



Kuvio 5: Green Wolf -panimon alavirran jakelukartta (Sourcemap 2020)

Alavirrasta tulisi pystyä kartoittamaan suorien asiakkaiden lisäksi myös vähintään virrat asiakkaiden asiakkaille ja pyrkiä reaaliaikaiseen informaation jakamiseen heidän kanssaan (Green 2001, 211; Mubarik ym. 2023, 2665). Alavirran toimitusketjukartoista tulisi myös ilmetä ne kohdat toimitusketjuissa, joissa vastuu kuljetettavasta tavarasta vaihtuu esimerkiksi toimitusehtojen mukaisesti (Fabbe-Costes ym. 2020, 1483).

3 Luonnollisen kielen käsittely

3.1 Luonnollisen kielen käsittelyn määritelmä ja taustaa

Luonnollisella kielellä tarkoitetaan mitä tahansa ihmiskommunikaatioon kehittyntä kieltä. Sen käsittelyllä viitataan taas ihmisten ja tietokoneen väliseen kommunikointiin luonnollisen kielen keinoin. Luonnollisen kielen käsittelyn voidaan nähdä koostuvan kahdesta toiminnallisuudesta: luonnollisen kielen ymmärtämisestä, jonka tarkoituksena on siirtää ihmisten käyttämä kieli tietokoneen ymmärtämään muotoon, ja luonnollisen kielen tuottamisesta, jonka tarkoituksena on luoda tietokoneen ymmärryksen pohjalta ihmisten ymmärtämää luonnollisen kielen tekstiä. (Li ym. 2021, 570; Khurana ym. 2023, 3715–3720.)

Vahvasti luonnollisen kielen käsittelyn taustalla oleva lingvistiikka (Eisenstein 2019, 1; Nadkarni ym. 2011, 544) eli kielitiede syntyi nykymuotoiseksi tieteenalaksi 1800-luvulla (Aronoff & Rees-Miller 2017, 4). Luonnollisen kielen käsittely syntyi omana käsitteenään 1950-luvulla, kun samoihin aikoihin esiteltä tekoälyn käyttöä lingvistiikassa lähdettiin soveltamaan, jolloin syntyi tietokone-lingvistiikan tieteenala (Nadkarni ym. 2011, 544; Li ym. 2021, 569). Chomsky (1956) esittelemät keinot luonnollisten kielten kieliopin teoreettisen analyysiin loivat pohjaa luonnollisen kielen tekstin koneelliseen käsittelyyn. Aikaisimmat käyttökohteet olivat yksinkertaisissa koneellisissa käännohjelmassa (Nadkarni ym. 2011, 544; Khurana ym. 2023, 3720). 1980- ja 90-luvuilla luonnollisen kielen käsittelyssä siirryttiin koneoppimiseen perustuvaan tilastolliseen kielenkäsittelyyn, jossa algoritmeille syötetään suuria määriä luonnollisen kielen ilmaisuja niiden selityksineen. Datan avulla kone oppii ymmärtämään luonnollisen kielen tekstiä (Nadkarni ym. 2011, 544; Khurana ym. 2023, 3720–3721.) 2000-luvulla luonnollisen kielen käsittelyyn on kehitetty neuroverkkoihin ja syväoppimiseen perustuvia ratkaisuja (Khurana ym. 2023, 3721–3722; Wichmann ym. 2020; Deng & Liu 2018, 6–10), joiden kyky käsitellä tekstiä on tilastollista kielenkäsittelyä merkittävästi parempi (Janiesch ym. 2021, 688). Aiemmin luonnollisen kielen käsittelyn tutkimus on keskittynyt lähinnä laskennallisen lingvistiikan tieteenalalle mutta tekstimuotoisen datan määrän kasvaessa myös luonnollisen kielen käsittelyn kehitys on kiihtynyt sen käyttökohteiden lisääntyessä (Li ym. 2021, 569).

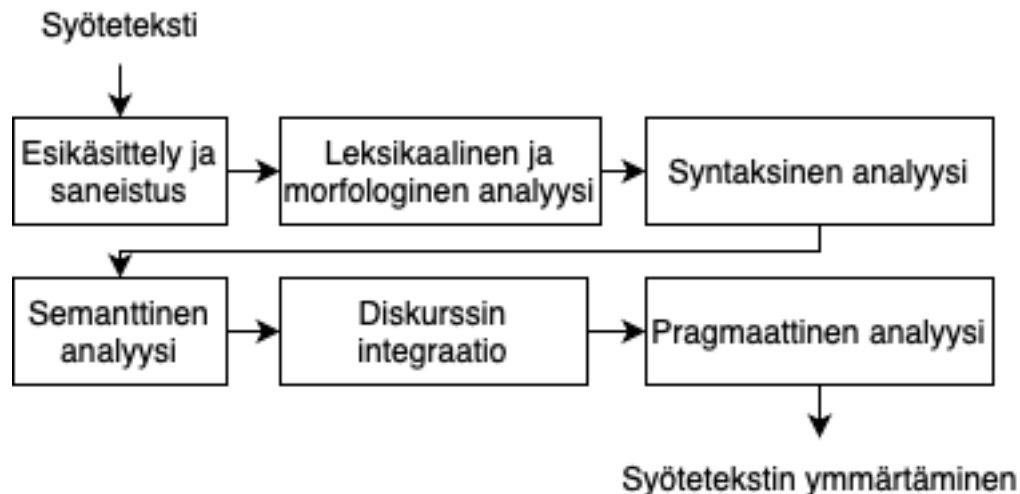
Lingvistiikan tutkimusalueista luonnollisen kielen käsittelyn kannalta olennaisia ovat fonologia, morfologia, syntaksi, semantiikka ja pragmatiikka (Khurana ym. 2023, 3715; Sharma ym. 2022, 259). Nämä muodostavat fonetiikan kanssa lingvistiikan päättutkimusalueet (Matthews 1991, 2–3; Pustejovsky & Stubbs 2012, 3–4). Fonologia tarkastelee eri kielten puheenmuodostusta (Katamba 1989, 1). Toimitusketjujen kartoittaminen luonnollisen kielen käsittelyn avulla kuitenkin perustuu tekstimuotoisiin datalähteisiin (Wichmann ym. 2018; Wichmann ym. 2020), joten fonologia ei ole kartoittamisen kannalta olennainen tutkimuskohde. Morfologia tarkastelee yksittäisten sanojen sisäistä rakennetta. Morfologiassa sanat jaotellaan yksittäisiksi morfeemeiksi, jotka edustavat yksittäistä sanan osaa. Sanojen sijamuodot määrittelevät kirjainyhdistelmät ovat esimerkki morfeemista. (Khurana ym. 2023, 3716; Matthews 1991, 1–9; Pustejovsky & Stubbs 2012, 3.) Morfologiaan liittyy vahvasti leksikologia, joka tarkastelee yksittäisten sanojen merkitystä ja niiden kielioppia. Leksikologia on leksikologian perusyksikkö, joka viittaa sanojen perusmuotoon. (Halliday & Yallop 2007, 4; Matthews 1991, 26; Pustejovsky & Stubbs 2012, 3.) Syntaksi tarkastelee lauseiden rakennetta ja merkityserojen muutosta rakenteiden muuttuessa. Lauseiden sanajärjestys on yksi esimerkki syntaksin tarkastelukohteista. (Khurana ym. 2023, 3717; Matthews 1991, 2; Pustejovsky & Stubbs 2012, 3.) Semantiikka eroaa syntaksista siten, että siinä tarkastellaan lauseiden sananmukaisen merkityksen sijasta niiden todellista merkitystä (Khurana ym. 2023, 3717; Matthews 1991, 2; Pustejovsky & Stubbs 2012, 3; Saeed 2016, 1). Semantiikkaan ja syntaksiin liittyy myös eri lauseiden välinen diskurssi tekstin sisällä. Esimerkiksi sana ”hän” viittaa yleensä kahden erillisen lauseen väliseen diskurssiin. (Khurana ym. 2023, 3717–3719; Pustejovsky & Stubbs 2012, 3.) Pragmatiikan ja semantiikan välille on vaikea määrittää selkeää eroa mutta pragmatiikka tarkastelee erityisesti tekstin käyttökontekstin vaikutusta sen ymmärtämiseen ja todelliseen merkitykseen (Khurana ym. 2023, 3718–3719; Matthews 1991, 2; Saeed 2016, 33; Pustejovsky & Stubbs 2012, 4). Pragmatiikan voidaan ymmärtää olevan lukijan tai kuulijan tulkintaa tekstin todellisesta merkityksestä (Saeed 2016, 33).

Luonnollisen kielen käsittelyn toteuttamiseksi sen taustalle tarvitaan laajoja tekstilähteitä. Yksi tärkeimmistä datan lähteistä ovat korpuukset, jotka ovat joko teksteistä tai puheesta johdettuja lingvistisen datan lähteitä. Luonnollisen kielen käsittelyllä automatisoitu toimitusketjujen kartoittamisen prosessi perustuu korpuksiin, jotka ovat koottu verkosta julkisesti saatavilla olevista tekstilähteistä, kuten blogeista ja

uutisista (Wichmann ym. 2018, 1729; Wichmann ym. 2020, 5321). Korpukset voivat sisältää tekstiin liittyviä annotaatioita tai ne voivat sisältää vain raakamuotoista tekstidataa (Khurana ym. 2023, 3730). Annotaatiot sisältävät teksteihin liittyvää metadataa, jotka ovat kasvavassa määrin automaattisesti luotuja. Annotaatiot saattavat sisältää tietoa esimerkiksi lauseen sanojen välisistä suhteista. (Grosman ym. 2020 2–3; Khurana ym. 2023, 3730; Pustejovsky ym. 2017, 1–4.) Annotoidut korpukset ovat tärkeitä datalähteitä erityisesti luonnollisen kielen käsittelyn ohjelmien kouluttamisessa (Banko & Brill 2001, 26; Khurana ym. 2023, 3727; Müller & Strube 2006, 198; Pustejovsky & Stubbs 2012, 1; Zhou & He 2008, 1114). Korpuksien annotaatiot saattavat olla yleisluontoisia mutta ne voivat liittyä myös tiettyyn lingvistiikan aihepiiriin. Annotoitujen korpuksien avulla voidaan luonnollisen kielen käsittelyn vaiheessa keskittyä syvemmin johonkin tiettyyn lingvistiikan aihepiiriin. (Müller & Strube 2006, 198; Newman & Cox 2020, 26.) Korpuksien annotointi aiheuttaa kuitenkin merkittävästi kustannuksia, mikä vähentää annotoitujen korpuksien sisältämää sanamäärä ja tekee niistä maksullisia (Banko & Brill 2001, 26; Grosman ym. 2020, 2; Orasan ym. 2007, 68; Snow ym. 2008, 254; Thieberger ym. 2016, 2). Raakamuotoista tekstiä sisältävät korpukset ovat sanamäärältään merkittävästi suurempia kuin annotoidut korpukset ja ne ovat usein ilmaiseksi saatavilla (Hiltunen ym. 2017; Säily & Tyrkkö 2021, 1–2). Erityisesti toimitusketjujen kartoittamisen kontekstissa hyödyllisen nettisivuilla sijaitsevan tekstuaalisen datan määrä on kasvanut merkittävästi (Grosman ym. 2020, 1; Li ym. 2021, 569; Wichmann ym. 2020, 5320). Yritysten kannalta mahdollisesti hyödyllisestä liiketoimintatiedosta merkittävä osa onkin tallennettuna rakenteettomassa muodossa, jota raakatekstiä sisältävät korpukset edustavat (Gharehchopogh & Khalifelu 2011, 2; Grosman ym. 2020, 1). Luonnollisen kielen käsittelyn ohjelmaa kehitettäessä tulee käyttää niin laadukkaasti annotoituja kuin myös isoja raakamuotoisia korpuksia, koska luotujen annotaatioiden laatu ja tekstilähteiden koko vaikuttavat luonnollisen kielen käsittelyn ohjelman suorituskykyyn (Grosman ym. 2020, 1–2; Rutowski ym. 2022, 3345; Wichmann ym. 2020, 5326).

3.2 Luonnollisen kielen käsittelyn prosessi

Luonnollisen kielen käsittelyn prosessi koostuu kuudesta päävaiheesta, joiden avulla tietokone pyrkii ymmärtämään luonnollisen kielen tekstiä (Kuvio 6).



Kuvio 6: Luonnollisen kielen käsittelyn prosessi (Dale 2010, 4; Li ym. 2021, 570; Sharma ym. 2022, 259)

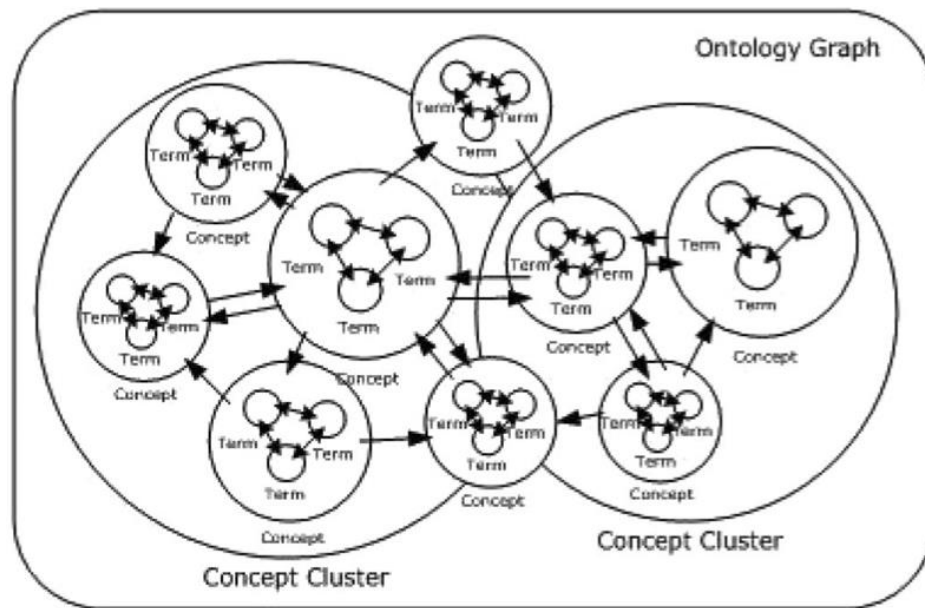
Prosessi alkaa syötetekstin esikäsittelystä ja saneistuksesta, joiden tarkoituksena on muuttaa teksti tietokoneelle helpommin ymmärrettävään muotoon. Esikäsittelyssä tekstistä poistetaan ylimääräiset merkinnät, kuten välimerkit ja tekstin ymmärtämisen kannalta epäolennaiset sanat. Saneistuksessa muokattu teksti jaetaan saneiksi, jotka edustavat yhtä sanaa. Sanejen muodostamat lauseet myös erotellaan toisistaan. Haasteen tekstin käsittelyyn ja selkeään määrittelyyn tuo erityisesti kielten erilaiset kirjoitusasut. Ensimmäisen vaiheen tuloksena on merkkijono, jossa alkuperäisen lauseen sanat ja lauseet ovat tietokoneelle ymmärrettävässä muodossa eroteltuna toisistaan. (Grefenstette & Tapanainen 1994, 2–4; Palmer 2010, 9–15; Tabassum & Patil 2020, 4864–4865.)

Tekstin esikäsittelyn jälkeen tietokone voi alkaa analysoidaan tekstiä. Ensimmäisenä tekstile tehdään leksikaalinen ja morfologinen analyysi, jossa tarkoituksena on muuttaa teksti toisistaan merkitykseltä eroaviin lekseemeihin. Yksi tekniikka tähän on esimerkiksi sanarunkohaku, jonka tarkoituksena on poistaa sanoista niiden päätteet ja muuttaa ne niiden perusmuotoon (Sharma ym. 2022, 258–259; Singh & Gupta 2017, 1). Toisena tekstile tehdään syntaktinen analyysi, jossa tarkoituksena on tarkastella lauseiden sanojen välisiä suhteita ja näiden avulla päätellä lauseiden merkityksiä. Jos lauseiden rakenne on virheellinen esimerkiksi sanajärjestyksen takia, lause hylätään merkityksettömänä, eikä sen analysointia jatketa tulevissa vaiheissa (Sharma ym. 2022, 259; Traxler 2017, 353.) Kolmantena tekstile tehdään semanttinen analyysi, jossa tarkoituksena ymmärtää lauseiden todellista merkitystä (Sharma ym. 2022, 259).

Neljäntenä tekstin sisäistä diskurssia eri lauseiden välillä analysoidaan ja pyritään löytämään niiden väliltä yhteyksiä. Tarkoituksena on ymmärtää yksittäisten lauseiden merkitys ottaen huomioon niiden diskurssi tekstin muiden lauseiden kanssa. (Sharma ym. 2022, 259). Viimeisenä tekstille tehdään pragmaattinen analyysi, jossa tarkoituksena on pyrkiä ymmärtämään konteksti, jossa tekstiä käytetään ja sen vaikutusta tekstin todelliseen merkitykseen (Sharma ym. 2022, 259; Li ym. 2021, 570). Tämän prosessin tuloksena tietokone on saanut jonkinnäköisen ymmärryksen tekstin sisällöstä ja pystyy sen perusteella tekemään mahdollisia jatkotoimenpiteitä.

3.3 Luonnollisen kielen käsittelyn jälkeinen tiedon visualisointi

Luonnollisen kielen käsittelyn prosessissa tekstin ymmärtämisen pohjalta luotu tieto muodostaa ontologian. Tietotekniikan kontekstissa ontologialla tarkoitetaan tarkkaan määriteltä tiedon käsitteellistämistä. Ontologia viittaa yksittäiseen olemassa olevaan tietoon jostain kokonaisuudesta, joka on tarkkaan määriteltä ja sen tiedetään olevan olemassa. Ontologiat koostuvat erilaisista alikokonaisuuksista, jotka liittyvät ontologian suurempaan kokonaisuuteen. Kun eri alikokonaisuuksien välille luodaan yhteyksiä, ne muodostavat tietokoneohjelman muodostaman tiedon verkostomaisen kokonaisuuden, jonka tietokone esittää ihmisen ymmärtämässä muodossa (Gruber 1993, 199; Grosman ym. 2020, 2.) Luonnollisen kielen käsittelyn avulla tekstimuotoisesta datasta voidaan muodostaa ontologioita ja niiden muodostamia kokonaisuuksia (Maedche & Staab 2001, 73).



Kuvio 7: Ontologiakartan hahmotelma (Liu ym. 2013, 118)

Toimitusketjujen visualisointi perustuu tekstin perusteella luotuihin kokonaisuuksiin, jotka edustavat toimitusketjun toimijoita ja niiden välisiä suhteita. Tämä kokonaisuus käytännössä kuvaa organisaation toimitusketjuverkostoa. Kokonaisuus voidaan visualisoida verkostomaisena ontologiakarttana (Kuvio 7), jonka pohjalta voidaan luoda toimitusketjukartta. (Wichmann ym. 2020, 5324.) Selkeästi määritellyn ontologiarakenteen etu on se, että sen sisältämä tieto voidaan yhdistää muissa tietokannoissa sijaitsevaan tietoon (Grosman ym. 2020, 2). Tällöin ontologiakartan avulla luotuun toimitusketjukarttaan voidaan lisätä toimijoihin liittyvää hyödyllistä lisätietoa muualta organisaation tietokannoista. Jotta toimitusketjukarttaan ei kuitenkaan päädy päällekkäisiä toimijoita ja suhteita, jokaisen toimijan ja luonnollisen kielen käsittelyn avulla löydetyn yksittäisen suhteen tulee esiintyä ontologiakartan visualisoinnissa vain kerran (Wichmann ym. 2020, 5333; Wichmann 2022, 126).

4 Luonnollisen kielen käsittelyn hyödyntäminen toimitusketjujen kartoittamisessa

Luonnollisen kielen käsittelyllä on havaittu olevan potentiaalia automatisoimaan toimitusketjujen kartoittamisprosessi mutta Wichmann ym. (2020) ja Wichmann (2022) kehittämät keinot eivät pysty vielä monimutkaisen kartan luontiin (Schöpfer & Kersten 2021, 80–82). Horisontaaliseen eli kilpailijoiden kartoittamiseen pyrkiviä luonnollisen kielen käsittelyn menetelmiä on tutkittu enemmän kuin vertikaaliseen toimittajien ja asiakkaiden kartoittamiseen pyrkiviä menetelmiä on tutkittu vähemmän (Schöpfer & Kersten 2021, 78–82). Tässä luvussa esitellään yksinkertaisen toteutuksen ominaisuuksia ja tuodaan myös esille ominaisuuksia, joita tulee kehittää tulevaisuudessa, jotta monimutkaisempien karttojen luominen ja teknologian laajempi käyttöönotto on tulevaisuudessa mahdollista.

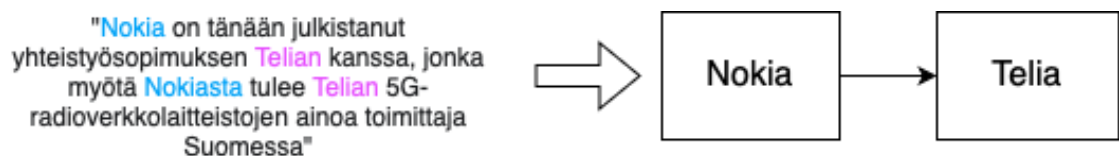
4.1 Toimitusketjujen kartoittamiseen soveltuvat luonnollisen kielen käsittelyn menetelmät

Toimitusketjujen kartoittamisessa luonnollisen kielen käsittelyä käytetään erityisesti tiedonpoimintaan julkisista tekstilähteistä (Wichmann ym. 2020, 5324). Tiedonpoiminta (eng. information extraction) on yksi luonnollisen kielen käsittelyn käyttökohteista, joka perustuu halutun tiedon löytämiseen rakenteettomasta tekstuaalisesta datasta luonnollisen kielen käsittelyn avulla ja sen muuttamista rakenteelliseen muotoon (Cowie & Lehnert 1996, 81; Piskorski & Yangarber 2013, 24). Tiedonpoiminta voidaan jakaa neljään alatoimintoon: nimien tunnistukseen (eng. named-entity recognition), yhteyksien poimintaan (eng. relation extraction), tapahtumien poimintaan (eng. event extraction) ja samaviitteisyyden tunnistamiseen (eng. co-reference resolution) (Piskorski & Yangarber 2013, 26–27; Tuo & Yang 2023, 7391).

Nimien tunnistamisessa tekstistä pyritään tunnistamaan kaikki sellaiset entiteetit, jotka ovat relevantteja datalähteestä poimittavan tiedon kannalta (Nadeau & Sekine 2007, 3; Piskorski & Yangarber 2013, 26; Wichmann 2022, 34). Toimitusketjujen kartoittamisessa nimien tunnistuksen alatoiminto nimien merkityksen selkeyttäminen (eng. named-entity disambiguation) on myös tärkeää, jotta luonnollisen kielen käsittelyn ohjelma tunnistaa tietyn yrityksen, johon tekstissä viitataan. Yksittäinen sana voi viitata moneen entiteettiin tekstissä, esimerkiksi sana Delta voi viitata useampaan yrityksiin,

kuten Delta Airlines tai Delta Faucet. Nimien merkityksen selkeyttäminen pyrkii määrittämään tietyn yksittäisen entiteetin, johon tekstissä viitataan. (Kandasamy & Cherukuri 2020, 1–2; Wichmann ym. 2022, 35.) Toimijoiden nimet saattavat myös esiintyä tekstissä tuotteiden nimien kanssa. Esimerkiksi sanapari ”Apple Iphone” viittaa Applen valmistamaan Iphone -puhelimeen, ei yritykseen kokonaisuudessaan. Nimien tunnistuksen tulee tunnistaa näiden mainintojen olevan tuotteita, eikä sekoittaa niitä toimijaentiteetteihin. (Wichmann 2022, 138.) Toimitusketjujen kontekstissa relevantteja entiteettejä ovat toimitusketjussa toimivat organisaatiot.

Yhteyksien poimimisessa pyritään löytämään yhteyksiä nimien tunnistuksessa löydettyjen entiteettien välille (Piskorski & Yangarber 2013, 27; Smirnova & Cudré-Mauroux 2019, 1; Tuo & Yang 2023, 7391). Toimitusketjujen kontekstissa organisaatioentiteettien välille pyritään löytämään yhteyksiä, jotka edustavat ostaja-toimittajasuhteita. Yksittäiset yhteydet muodostavat yhdessä ontologian kokonaisuuden ja näin pohjan toimitusketjun verkoston kokoamiseen. Suurena ongelmana yhteyksien luomisessa on kuitenkin se, että tällä hetkellä nimenomaisesti ostaja-toimittajasuhteiden määrittämiseen tarkoitettuja nimien tunnistamisen ja yhteyksien poiminnan keinoja ei ole. (Wichmann ym. 2020, 5325; Wichmann 2022, 6, 95, 173–174.) Entiteettien kaltaisesti yksittäisiä sanoja voidaan käyttää implikoimaan monenkaltaisia suhteita. Sana ”deal” voi viitata esimerkiksi toimittajasopimukseen tai toisen yrityksen toimintojen hankkimiseen. (Wichmann 2022, 138.)



Kuvio 8: Esimerkki tekstistä löydetystä ostaja-toimittajasuhteesta

Näitä kahta tiedonpoiminnan keinoa hyödynnetään Wichmann ym. (2020) ja Wichmann (2022) esittelemässä yksinkertaisessa kartoittamisessa mutta tulevaisuuden toteutuksien tulee myös hyödyntää samaviitteisyyden tunnistamista ja tapahtumien poimimista tiedon poiminnassa. Samaviitteisyyden tunnistamisessa pyritään tunnistamaan, jos sama nimien tunnistamisessa löydetty entiteetti esiintyy useamman kerran samassa tekstilähteessä (Piskorski & Yangarber 2013, 26–27). Samaviitteisyyden tunnistamiseen liittyy vahvasti luonnollisen kielen käsittelyn diskurssin integraatio (Ledoux ym. 2007, 801). Esimerkiksi jos tekstilähteessä esiintyy sanapari ”the company”,

samaviitteisyyden tunnistamisella tietokone pyrkii tunnistamaan, mihin tekstilähteessä esiintyvään toimijaentiteettiin sanapari viittaa. Samaviitteisyyden tunnistamisen hyödyntäminen parantaisi yhteyksien poimintaa, koska yksittäisessä lauseessa saattaa esiintyä yhdessä viittaus aiemmin esiintyneeseen toimijaentiteettiin ja maininta uudesta toimijaentiteetistä. Ilman samaviitteisyyden tunnistamista tietokoneen on vaikea tunnistaa tarkoitettua suhdetta tekstissä. (Wichmann 2022, 124.)

Tapahtumien poimimisessa pyritään löytämään toisistaan uniikkeja tapahtumia. Tiedonpoiminnassa tapahtumalla tarkoitetaan uniikkia tapahtumaa, joka johtaa johonkin muutokseen ja johon liittyy nimien tunnistuksessa löydettyjä entiteettejä. (Piskorski & Yangarber 2013, 27; Xiang & Wang 2019, 173111.) Toimitusketjujen kontekstissa esimerkiksi ostaja-toimittajasuhteen alkaminen tai loppuminen kahden organisaation välillä on esimerkki tiedonpoiminnan tapahtumasta. Tapahtumanpoiminnan merkitys on siis merkittävä toimitusketjukartan päivitettävyyden kannalta, koska sen avulla pystytään havaitsemaan, jos toimitusketjukartan toimijoiden välisissä suhteissa tapahtuu muutoksia.

Toimitusketjujen kartoittamista varten tulee luoda tiedonpoimintaa hyödyntävä luokittelijana toimiva koneoppiva ohjelma, jonka toiminta perustuu moniluokkaiseen luokitteluun (Wichmann ym. 2020, 5325; Wichmann 2022, 100). Koneoppivalla ohjelmalla tarkoitetaan tietokoneohjelmaa, joka pyrkii oppimaan datalähteiden kokonaisuuksien tunnistamista ja niiden välisiä suhteita (Janiesch ym. 2021, 685). Moniluokkaisessa luokittelussa ohjelmalle on määritelty joukko toisiaan poissulkevia luokkia. Jokainen tiedonpoiminnassa tarkasteltu tekstilähde kuuluu johonkin näistä luokista. Entiteettien luokkia kuitenkin yhdistää yleisesti jokin selkeä yläluokka. (Denis ym. 2024, 2; Wichmann ym. 2020, 5325–5327.) Yksinkertaisessa toimitusketjujen kartoittamisessa, jossa toimijoiden välille luodaan ainoastaan suhteita, toimijasuhteiden luokat voidaan jakaa viiteen tai kuuteen erilaiseen luokkaan, joista luokat A-C edustavat sellaisia tekstilähteitä, joista löytyy toimittajasuhteita ja luokat D-F sellaisia tekstilähteitä, joista toimittajasuhteita ei löydy.

Toimittajasuhteiden luokat voivat kuvata suoran toimittajasuhteen tilannetta A, jossa toimija A toimittaa toimijalle B, käänteisen toimitussuhteen tilannetta B, jossa toimija B on toimijan A toimittaja tai suunnitteilla olevan toimitussuhteen tilannetta C, jossa toimija A ja toimija B neuvottelevat mahdollisesta toimittajasuhteesta. Koska

toimitusketjukartassa toimijoiden välisten virtojen suunta on tärkeä tieto, luokissa on tärkeä tulla toimijoiden välisten virtojen suunta esille. Toimittajasuhteen puuttumista kuvaavat luokat voivat kuvata omistussuhteen tilannetta D, jossa toimija A omistaa osan tai kokonaan toimijan B toiminnan, koneellisessa nimien tunnistuksessa virheellisesti määriteltyjen organisaatioiden tilannetta E, jossa esikäsittelyn nimien tunnistus oli virheellisesti päästänyt läpi tekstilähteen manuaaliseen annotointiin tai tilannetta F, jossa tekstilähteestä ei löydy toimijoiden välistä toimittajasuhdetta. Jos luokittelua halutaan yksinkertaistaa, tilanteet E ja F voidaan yhdistää yhteiseksi luokakseen. Tilanne D tarvitaan erilliseksi luokakseen, koska muuten ohjelma saattaisi tulkita esimerkiksi tekstilähteen, jossa kerrotaan toimijan A ostavan toimijan B toiminnan, kuuluvan tilanteen B luokkaan. Tilanteen D luokka ei siis implikoi toimittajasuhdetta mutta sen avulla voidaan suodattaa toimittajasuhteista pois sellaiset tilanteet, joissa kyseessä on toimittajasuhteen sijasta omistajuussuhde. (Wichmann ym. 2020, 5327–5328; Wichmann 2022, 96, 99, 103–105.)

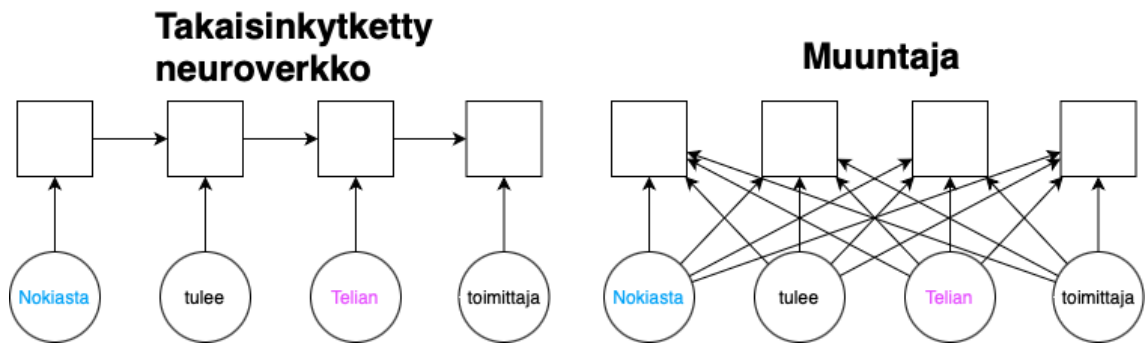
Käytännössä tieto entiteeteistä ja niiden välisen suhteen luokasta tallennetaan monikkomuotoiseen tietorakenteeseen (Wichmann ym. 2020, 5325; Wichmann 2022, 96). Luokkien määrä vaikuttaa tiedonpoimintaan käytettävän luokittelijan suorituskykyyn ja sen käytön luomiin vaatimuksiin. Suurempi luokkien määrä mahdollistaa tarkempien semanttisten erojen tunnistamisen tekstilähteistä mutta vaatii myös enemmän annotoituja tekstilähteitä, joita pitää määrittää kaikille luokille hyvin paljon. Pienempi luokkien määrä vaatii vähemmän annotointeja mutta luokittelijan kyky erotella entiteettejä toisistaan heikkenee. Mahdollisimman tarkka toimitusketjujen kartoittaminen vaatii mahdollisimman paljon luokkia. (Wichmann ym. 2020, 5332–5333.) Tarkkuutta saadaan lisättyä myös lisäämällä monikkoon lisättävän tiedon määrää. Tietoa voidaan lisätä esimerkiksi suhteen maantieteellisestä jakautumisesta, toimijoiden välisessä virrassa kulkevista asioista tai suhteen kriittisyydestä. Muuttujien lisääminen kuitenkin monimutkaistaa luokittelua ja luokittelijan luontia. Ominaisuuksien lisääminen tulevaisuudessa on kuitenkin välttämätöntä, jotta luonnollisen kielen käsittelyä voidaan hyödyntää yritysmaailmassa toimitusketjujen kartoittamiseen.

Luokittelijan toiminta perustuu tiedonpoiminnassa siihen, että luokittelija pyrkii etsimään tekstilähteistä luokille ominaisia piirteitä ja määrittelee raakamuotoiselle tekstilähteelle todennäköisimmän luokan, johon se kuuluu (Wichmann ym. 2020, 5329).

Luokittelijan tulee perustua sekventiaalisen datan käsittelyyn soveltuvaan koneoppivan ohjelman toteutukseen, koska luonnollinen kieli edustaa sekventiaalista dataa, jossa datan sisäisellä järjestyksellä on merkitystä. Luonnollisen kielen sekventiaalisuus ilmenee esimerkiksi lausejärjestyksen merkityksellä. (Lyon & Frank 1998, 198.) Wichmann ym. (2020) ja Wichmann (2022) luomien luokittelijoiden toiminta perustuu sekventiaalisen datan käsittelyyn soveltuvaan kaksisuuntaisen takaisinkytketyn neuroverkon (eng. bi-directional recurrent neural network) toimintaan.

Takaisinkytkettyjen neuroverkkojen käytössä haasteena kuitenkin on häviävän gradientin ongelma (eng. vanishing gradient problem), jossa tekstilähteen koon kasvaessa tekstilähteen loppupäässä sijaitsevat sanat vaikuttavat yhä vähemmän neuroverkon tekemän laskennan lopputulokseen (Roodschild ym. 2020, 351). Piirteiden etsinnässä tietyn luokalle ominaisten sanojen esiintymisen sanajärjestystä suurempi merkitys on sillä, että luokalle ominaiset sanat ylipäättään esiintyvät tekstilähteissä. (Wichmann ym. 2020, 5333). Tällöin toimijasuhteelle merkittävät sanat saattavat jäädä vähälle painoarvolle, jos ne sijaitsevat pitkän tekstilähteen loppupäässä, jolloin tekstilähteen sisältämä toimijasuhde saattaa jäädä luokittelijalta havaitsematta. Kaksisuuntaista takaisinkytkettyä neuroverkkoa voidaan käyttää häviävän gradientin ongelman osittaisessa ratkaisemisessa (Schuster & Paliwal 1998, 2675) mutta kehittyneempiä ratkaisuja on jo saatavilla.

Muuntajat (eng. transformer) pystyvät takaisinkytkettyjä neuroverkkoja paremmin käsittelemään pitkiä tekstilähteitä (Vaswani ym. 2017, 6000–6001). Pidempien tekstilähteiden hyödyntäminen perustuu siihen, ettei muuntajan tarvitse takaisinkytketyn neuroverkon tapaan käsitellä tekstilähteitä sanajärjestyksen mukaisesti, mikä mahdollistaa rinnakkaislaskennan tekstilähteen käsittelyssä (Kuvio 9; Zaheer ym. 2020, 17283). Kun pidempien tekstilähteiden käsittely paranee, todennäköisyys toimijasuhteen kannalta olennaisten sanojen painoarvon häviämisestä pienenee. Tulevaisuuden toteutuksissa tulisi siis tarkastella muuntajan toimintaan perustuvaa luokittelijaa (Wichmann 2022, 215). Toteutuksesta riippumatta luokittelija ennustaa tekstilähteen sisältämän datan perusteella, minkä luokan entiteettejä tekstilähde sisältää (Wichmann ym. 2020, 5329–5330).



Kuvio 9: Takaisinkytketyn neuroverkon ja muuntajan toiminnan hahmotelma

Toimitusketjujen kasvava globaalius (Christopher & Towill 2002, 1) tekee datanlähteistä monikielisiä, koska toimitusketjun tiettyä osaa käsittelevää dataa löytyy eniten maantieteellisesti paikallisen alueen kielellä. Luodun luokittelijan tulisi siis pystyä tulkitsemaan datalähteitä monella kielellä. Tämä on kuitenkin erityinen haaste, koska monikielisyys opettaminen kielimallille vaatii myös monikielistä koulutusaineistoa (Litschko ym. 2022, 152), mikä monimutkaistaa tiedonhankintaa ja koulutusaineiston annotointia. Monikielisyys on yksi selkeä kehityskohde luokittelijan toiminnassa, koska Wichmann ym. (2020) ja Wichmann (2022) kehittämät luokittelijat pystyvät hyödyntämään ainoastaan englanninkielisiä tekstilähteitä.

4.2 Haasteet luonnollisen kielen käsittelyyn perustuvassa toimitusketjujen kartoittamisessa

Wichmann ym. (2018), Wichmann ym. (2020) ja Wichmann (2022) tuovat esille haasteita, joita toimitusketjujen kartoittamiseen tarkoitetun luonnollisen kielen käsittelyn ohjelman tulee pystyä ratkaisemaan. Nykyiset toteutukset eivät pysty vielä vastaamaan läheskään kaikkiin haasteisiin. Haasteet voidaan jakaa kahteen: toimitusketjujen monimutkaisesta luonteesta johtuviin ongelmiin ja rajallisesta tiedon saatavuudesta johtuviin ongelmiin.

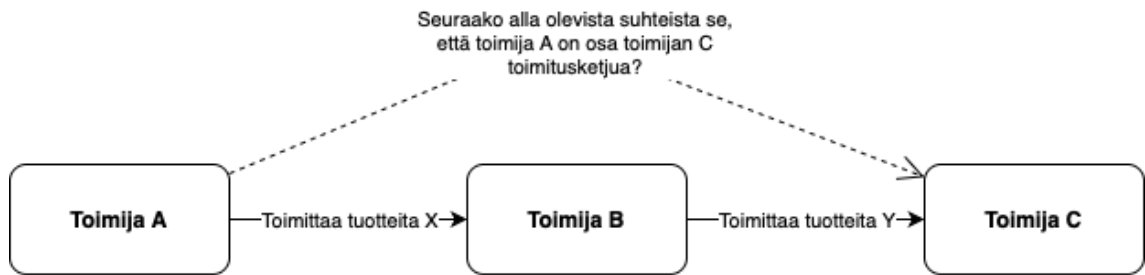
4.2.1 Toimitusketjujen monimutkaisesta luonteesta aiheutuvat haasteet

Joissain tapauksessa kaksi toimitusketjun toimijaa voivat samanaikaisesti toimia toisilleen toimittajana ja asiakkaana. Tämän takia tiedonpoiminnassa tulee erotella nämä tilanteet toisistaan esimerkiksi määrittämällä suhteen luonteen määrittävä tuote.

Yksittäisellä toimijalla saattaa siis olla toimitusketjussa monta eri roolia (Brintrup 2010, 637), jolloin se saattaa toimia monella eri toimitusketjun tasolla. Tällaisessa tilanteessa

luokittelijan saattaa olla vaikea päätellä, mihin toimitusketjun tasolle toimija oikeasti sijoitetaan. Toimijaentiteeteistä tulisi myös tunnistaa, liittyykö suhde organisaation päätoimintoon vai tukitoimintoon ja sivuuttaa tukitoiminnoiksi tunnistetut suhteet. Toimitusketjukartta sisältää ensisijaisesti tietoa vain toimijoista, jotka liittyvät ensisijaisiin toimintoihin. Luokittelijan tulee myös pystyä tunnistamaan logistiikkatoimijat erityiseksi toimijaksi toimitusketjussa. Luokittelija saattaa helposti tunnistaa logistiikkatoimijan virheellisesti toimittajaksi tai asiakkaaksi. Yksittäiset valtiot saattavat myös olla toimitusketjun toimijoita, erityisesti puolustussektorilla (Gansler ym. 2004, 4–5). Nimien tunnistamisessa valtiot saatetaan kuitenkin helposti tunnistaa toimijan sijasta maantieteelliseksi entiteetiksi.

Toimijoiden välisten suhteiden muodostamiseen liittyy myös olennaisesti matematiikan loogiseen päättelyyn liittyvä transitiivisuus ja siihen liittyvät ongelmat. Transitiivisuus tarkoittaa kykyä päätellä kahden erillisen entiteetin suhde aiemmin todistettujen väitteiden avulla (Goodwin & Johnson-Laird 2008, 321–322). Toimijoiden välisiä suhteita muodostaessa ongelma liittyy siihen, voiko luokittelija päätellä todistamiensa väitteiden $A = \text{Toimija } A \text{ on toimija } B:n \text{ toimittaja}$ ja $B = \text{Toimija } B \text{ on toimija } C:n \text{ toimittaja}$ pohjalta todistaa väitteen $C = \text{Toimija } A \text{ kuuluu toimija } C:n \text{ toimitusketjuun}$. Tämä johtopäätös ei kuitenkaan läheskään aina ole tosi, koska toimija A saattaa toimittaa toimijalle B jotain tuotetta, joka ei liity toimijan C toimitusketjuun millään tavalla. Tällöin luonnollisen kielen käsittelyn ohjelman pitää pystyä erottelemaan toisitaan tilanteet, joissa väite on tosi tai epätosi. Yksi keino tähän on määrittää kaikille löydetyille suhteille lopputuote, johon ne liittyvät (Wichmann ym. 2020, 5334, Wichmann 2022, 87). Transitiivisuuden ongelman ratkaiseminen on yksi olennainen osa luonnollisen kielen käsittelyyn perustuvan toimitusketjujen kartoittamisen automatisoinnissa, koska se vaikeuttaa esimerkiksi toimijoiden sijoittamista oikealle toimitusketjun tasolle (Wichmann ym. 2020, 5325; Wichmann 2022, 137).



Kuvio 10: Transitiivisuuden ongelman hahmotelma (Wichmann ym. 2018, 1729, Wichmann 2022, 132)

Luokittelijaa luotaessa sille tulee määrittää kultaisen standardin data, jonka avulla sitä voidaan kouluttaa tunnistamaan tekstistä luokille tyypillisiä ominaisuuksia. Kultaisella standardilla tarkoitetaan tunnettua ja laajasti hyväksyttävää määritelmää jostain kokonaisuudesta (Cardoso ym. 2014, 28). Luonnollisen kielen käsittelyn kontekstissa korpuksien annotaatiot edustavat tekstilähteiden kultaisen standardin dataa (Grosman ym. 2020, 2–3). Kultaisen standardin datan määrittäminen on hyvin aika- ja resurssivaativaa, koska se tulee lähes aina toteuttaa käsin (Ittoo ym. 2016, 103). Vajavaisestakin datasta voidaan löytää entiteettejä, jos tarpeeksi monta kultaisen standardin ominaisuutta esiintyy tekstissä. Toimitusketjujen monimutkaisuus vaikeuttaa kultaisen standardin datan määrittämistä, jolloin sen määrittäminen voi olla kuitenkin ihmisillekin haastavaa (Evert ym. 2016, 24). Jos kultaisen standardin dataa on vaikea määrittää ihmistenkin toimesta, luonnollisen kielen käsittelyyn perustuvaa automaatiota on hyvin haastavaa toteuttaa.

Toimitusketjujen dynaamisuus tuo myös vaateita luonnollisen kielen käsittelyssä käytettävän datan ajankohtaisuudesta. Jos käytettävä data on vanhaa, sen pohjalta voidaan luoda toimijoiden välille suhteita, joita ei oikeasti ole enää olemassa. Suhde on voitu katkaista mutta on myös mahdollista, että toimija on muuttanut nimeään tai sen toiminta on ostettu, jolloin vanhojen tekstilähteiden pohjalta luotu toimijasuhde on virheellinen. Luokittelijalla luotuihin suhteisiin on hyvä lisätä käytetyille datalähteille aikaleima, jolloin vanhemmasta datasta löydettyjä toimittajasuhteita voidaan tarkastella tarkemmin manuaalisesti mahdollisten virheiden tunnistamiseksi. Aikaleiman avulla datan luotettavuudelle voidaan asettaa eräänlainen puoliintumisaika, mitä vanhempi lähde on, sitä epäluotettavampaa siitä saatava tieto on.

4.2.2 Rajallisesta tiedon saatavuudesta aiheutuvat haasteet

Toimitusketjujen dynaamisuuteen liittyy myös se, että toimijasuhteisiin liittyvät tekstilähteet ovat yleisesti positiivissävytteisiä ja sisältävät tietoa toimittajasuhteiden alkamisesta tai jatkumisesta. Negatiivissävytteisiä tekstilähteitä esimerkiksi toimittajasuhteiden päättymisestä löytyy vähemmän. Tämä vaikeuttaa erityisesti vähemmän tärkeiden toimittajasuhteiden päivittämistä, joiden kesto on strategisia toimittajasuhteita lyhyempiä (Dubois ym. 2021, 697; Li ym. 2022, 303).

Luvussa 2.3 mainittu toimittajadatan yksityisyys korostuu erityisesti luonnollisen kielen käsittelyyn perustuvaan toimitusketjujen kartoittamiseen, koska kaikki tarvittava data ei ole julkisesti saatavilla. Tekstilähteiden saatavuus vaihtelee sektoreittain, esimerkiksi puolustussektorilla datan saatavuus saattaa tuottaa vaikeuksia (Giacomello & Preka 2023, 534). Tämän takia luonnollisen kielen käsittelyyn perustuvan toimitusketjujen kartoittamisen soveltuvuus vaihtelee sektoreittain. Luokittelijan tulee siis pystyä tunnistamaan entiteettejä ja niiden välille yhteyksiä myös silloin, kun niihin liittyvä data on vajavaista. Verkossa sijaitsevan datan luotettavuus vaihtelee myös paljon (Popat ym. 2016, 2173), joten ohjelman tulisi pystyä arvioimaan käyttämiensä lähteiden luotettavuutta ja joko hylätä tai merkitä ne manuaalista arviointia varten, jos niiden luotettavuutta on syytä epäillä. Eri sektoreilla saatetaan myös käyttää sektorille ominaista kieltä, jota luokittelijan tulee ymmärtää. Tekstilähteissä samoihin toimijoihin saatetaan myös viitata eri nimillä ja luonnollisen kielen käsittelyn ohjelman tulee pystyä yhdistämään nämä maininnat yhteen entiteettiin, joihin kaikissa teksteissä viitataan. Esimerkiksi merkkijonot ”Intesa Sanpaolo S.p.A” ja ”Intesa San Paolo bank” viittaavat samaan organisaatioon, italialaiseen Intesa Sanpaolo -pankkikonserniin (Basile ym. 2024, 1).

Joissain luonnollisen kielen käsittelyn menetelmissä entiteettien toistuvuus viittaa niiden tärkeyteen (Juluru ym. 2021, 1424) ja toimittajasuhteet saattavat esiintyä tekstilähteissä usein, jos esimerkiksi myyntimäärät ovat suuria. Toimittajasuhteisiin liittyvissä tekstilähteissä erilaisten toimittajasuhteiden toistuva esiintyminen tekstilähteissä ei kuitenkaan välttämättä implikoi niiden tärkeyttä, koska riskienhallinnan näkökulmasta toimittajasuhde voi olla kriittinen, vaikkei toimijoiden väliset myyntimäärät olisi suuria (Simchi-Levi ym. 2014, 99, Wichmann 2022, 78). Luokittelijan tulee siis pystyä arvioimaan toimijan tärkeyttä muillakin kriteereillä kuin

entiteetin esiintymistoistuvuudella. Toimijoiden välisiä suhteita ei myöskään usein selkeästi ilmaista tekstilähteissä, minkä takia luokittelijan tulee pystyä yhdistämään erillisiä lauseita toisiinsa toimijasuhteiden päättelemiseen.

4.3 Vaatimukset yrityksille toimitusketjujen kartoittamisen automatisoimiseksi luonnollisen käsittelyn avulla

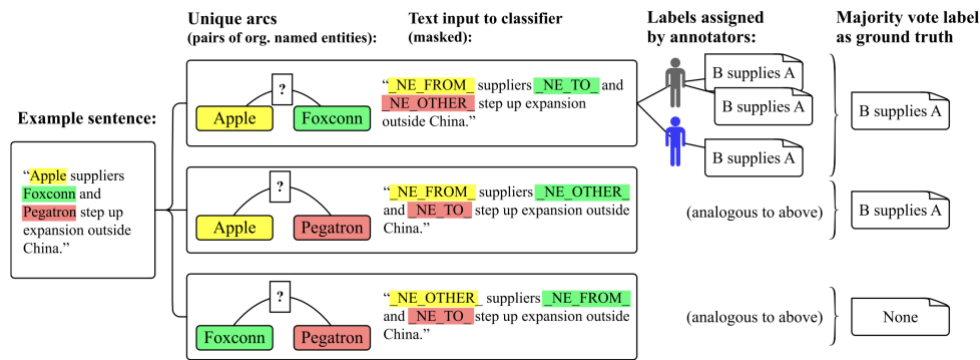
Jotta organisaatio voi automatisoida kartoittamisprosessinsa, organisaation tulee toteuttaa luvussa 4.1 esitetty luokittelijaohjelma käytännössä. Luokittelijaohjelma tarvitsee myös korpuksen, josta tekstilähteitä voidaan luokitella. Verkossa sijaitsevista tekstilähteistä koostuvat suuret korpuksat sisältävät yleisluontoisia tekstilähteitä, joita ei ole kategorisoitu millään tavalla (Hiltunen ym. 2017). Yleisluontoisia tekstilähteitä sisältävissä korpuksissa toimittajasuhteiden esiintyvyys on kuitenkin melko alhainen. Tällaisten lähteiden annotointi on resurssivaativuuteen verrattuna hyvin kannattamatonta, koska annotoitavasta tekstistä löytyy hyvin vähän esimerkkejä toimijoiden välisistä suhteista. (Wichmann ym. 2020, 5326.) Tämän takia, jotta luonnollisen kielen käsittelyn ohjelman käyttämät korpuksat olisivat tarpeeksi laajoja mutta sisältäisivät tarpeeksi toimitusketjukartalle relevanttia dataa, organisaation tulee itse kerätä tekstilähteitä ja luoda omaan käyttöönsä soveltuva korpus, josta voidaan hakea tietoa toimitusketjukarttaa varten (Wichmann 2022, 140, 166).

Korpukseseen voidaan hakea tekstilähteitä, jotka esimerkiksi sisältävät jonkun tietyn sanan, kuten tietyn tunnetun toimitusketjun toimijan nimen (Wichmann 2022, 165). Jos organisaatiolla ei ole resursseja tiedonkeruuseen, olemassa olevia korpuksia voidaan suodattaa hakusanojen kanssa, jotka liittyvät oman organisaation toimitusketjuun. Suodattamisenkin jälkeen toimittajasuhteiden esiintyvyys saattaa kuitenkin jäädä pieneksi. (Wichmann ym. 2020, 5326.) Osalle kerätyistä tekstilähteistä tulee myös lisätä annotaatioita, joiden avulla luokittelija ohjelma voi oppia luvussa 4.1 mainituille moniluokkaisen luokittelun luokille ominaiset piirteet. Nämä tekstiesimerkit voidaan esikäsitellä ja niistä voidaan etsiä toimijaentiteetit nimien tunnistuksen avulla koneellisesti, jotta annotaattoreiden tarvitsee vain määritellä vain esimerkin luokka ja sen ominaiset piirteet. Koska luokittelijan tarkoituksena on löytää toimittajasuhteita, koneellisessa nimien tunnistuksessa tulee löytää tekstilähteestä vähintään kaksi toimijaentiteettiä, jotta se voidaan lisätä manuaaliseen annotointiin. Nimien tunnistus on kannattavaa toteuttaa useamman ohjelmiston avulla, jotta nimien tunnistuksessa

vältytään mahdolliselta yksittäisen ohjelmiston aiheuttamilta virheellisiltä entiteettien tunnistuksilta.

Jos useampi nimentunnistusohjelma löytää samat toimijaentiteetit tekstilähteestä, tekstilähde voidaan lisätä manuaaliseen annotointiin. Annotoinneilla luodaan luokittelijaa varten tekstilähteistä positiivisia esimerkkejä, jotka edustavat luvussa 4.1 mainittuja luokkia A-C ja negatiivisia esimerkkejä, jotka edustavat luokkia D-F. Annotointien avulla ohjelma oppii tunnistamaan luokkien piirteitä raakamuotoisista tekstilähteistä. Ohjelman oppimisen kannalta positiiviset esimerkit ovat tärkeämpiä kuin negatiiviset esimerkit (Wichmann ym. 2020, 5326, Wichmann 2022, 106–107, 123–125.) Jos tekstiesimerkistä löytyy useampi kuin kaksi toimijaentiteettiä, jokaisesta entiteetistä tulee määrittää sen suhde esimerkissä esiintyviin muihin entiteetteihin (Kuvio 11; Wichmann ym. 2020, 5330). Kun luokittelija käy läpi annotoituja koulutuslähteitä, ohjelmalta piilotetaan lähteen sisältämät toimijoiden nimet, jotta ohjelma ei opi etsimään toimijasuhteita vain kyseisten toimijoiden välillä (Kuvio 11; Wichmann ym. 2020, 5328, Wichmann 2022, 109).

Annotoinnissa tulee kuitenkin välttää vain piirteiltään samankaltaisten esimerkkien luontia, koska muuten ohjelma oppii positiivisten esimerkkien edustavan suppeaa ja tarkkaa määriteltyä joukkoa, jolloin ohjelma etsii raakamuotoisista tekstilähteistä vain tietynlaisia toimijasuhteita, jolloin oikeasti toimijasuhteita kuvaavat entiteetit jäävät helposti huomaamatta (Wichmann ym. 2020, 5326–5327). Tätä ohjelman haun vinoumista tietynkaltaisiin piirteisiin kutsutaan koneharhaksi (eng. bias) (Mehrabi ym. 2021, 2). Jos positiivisten esimerkkien joukko on taas liian suuri, tapahtuu usein tiettyjen positiivisten esimerkkien piirteiden ylisovittamista (eng. overfitting), jos niiden esiintymistiheys on keskiarvoisia positiivisten esimerkkien piirteitä suurempaa. Ylisovittamisen seurauksena oppinut ohjelma saattaa löytää raakatekstistä vääriä positiivisia (eng. false positive), jos todellisesti negatiivisessa tekstilähteessä ylisovitetut piirteet esiintyvät toistuvasti, jolloin ohjelma tulkitsee tekstilähteen edustavan positiivista esimerkkiä. (Peng & Nagata 2020, 2; Wichmann 2020, 5327.)



Kuvio 11: Esimerkki yksittäisen tekstilähteen annotoinnista (Wichmann ym. 2020, 5329, Wichmann 2022, 111)

Luonnollisen kielen käsittelylle monimutkaisten toimitusketjujen luonteen (Luku 4.2) takia korpuksat ovat kannattavaa annotoida manuaalisesti toimitusketjujen toimintaan perehtyneiden asiantuntijoiden toimesta (Wichmann 2022, 149). Annotaatioprosessia varten organisaation tulee allokoita omasta työvoimastaan resursseja annotaatioprosessia varten tai ulkoistaa annotaatioprosessi ulkoiselle toimijalle. Toteutustavasta huolimatta annotaatioprosessissa on kannattavaa olla mukana projektinjohtaja, kuraattori ja useampia annotaattoreita (Biemann ym. 2017, 237–238). Projektinjohtaja vastaa yleisesti annotaatioprosessin etenemisestä ja tavoitteista, sekä annotaattoreiden kouluttamisesta (Biemann ym. 2017, 237–238). Kuraattorin tehtävänä on yhdistää eri annotaattoreiden korpukseseen tekemät annotaatiot yhteen dokumenttiin ja tarkistaa tehtyjen annotaatioiden paikkansapitävyys. Kuraattorin tehtävä on tärkeä annotaatioprosessin laadunvarmennuksen kannalta. (Biemann ym. 2020, 238; Grosman ym. 2020, 3.) Jos organisaatiolla ei ole resursseja kuraattorin käyttöön, tekstilähteiden luokaksi voidaan määrittää annotaattoreiden tekemien luokittelujen perusteella tekstilähteessä yleisimmin esiintyvä luokka (Wichmann ym. 2020, 5328, Wichmann 2022, 108). Annotaattoreiden tehtävänä on käydä läpi projektijohtajien antamia dokumentteja ja lisätä sinne annotaatioita projektinjohtajan ohjeiden mukaisesti (Biemann ym. 2017, 237; Wichmann ym. 2020, 5328). Sama tekstilähde on kannattavaa annotoida useamman annotaattorin toimesta (Wichmann 2022, 149).

Annotaatioiden laadun varmentamiseksi voidaan verrata annotaattoreiden tekemien annotaatioiden yhtenäisyyden lisäksi myös annotaattoreiden omien annotaatioiden yhtenäisyyttä antamalla annotaattoreille satunnaisesti samoja tekstilähteitä annotoitavaksi (Wichmann 2020, 5328; Wichmann 2022, 149). Annotaatioiden luontia varten prosessiin osallistuvilla tulee myös olla käytössään annotointiohjelmisto

(Biemann ym. 2017, 237; Finlayson & Erjavec 2017, 174; Wichmann ym. 2020, 5328). Jos kyseistä annotaatioprosessia varten ei ole soveltuvaa annotointiohjelmistoa, tulisi sekin kehittää organisaation toimesta (Finlayson & Erjavec 2017, 175–176).

4.4 Automaattisen kartoittamisen hyödyt verrattuna perinteiseen kartoittamiseen

Toimitusketjujen kartoittaminen voi olla hyvinkin monimutkainen prosessi, joka vaatii paljon työpanosta ja aiheuttaa näin kustannuksia ja vie paljon aikaa (Luku 2.3). Prosessin automatisoinnilla nopeutetaan prosessia ja voidaan saavuttaa merkittäviä kustannustehokkuuden nousun aiheuttamia säästöjä. Hyvin kehitetty kielimalli pystyy myös ymmärtämään useampia kieliä, jolloin hyödynnettävissä olevan datan määrä kasvaa merkittävästi verrattuna esimerkiksi toimitusketjukarttaa työstävän perinteisen työryhmän rajalliseen kielitaitoon. (Wichmann 2022, 209.) Automatisointi mahdollistaa myös perinteistä kartoittamista merkittävästi laajempien rakenteettomien datalähteiden käsittelemisen ja niiden sisältämän liiketoimintatiedon hyödyntämisen toimitusketjujen johtamisessa (Wichmann 2022, 213). Tämä on tärkeä lisäys kartoittamiseen, koska suurin osa hyödynnettävissä olevasta liiketoimintatiedosta on tallennettuna rakenteettomaan muotoon (Gharehchopogh & Khalifelu 2011, 2; Grosman ym. 2020, 1). Rakenteettoman datan hyödyntäminen perustuu luvussa 4.1 esiteltyyn tiedonpoiminnan menetelmään. Automatisoitu kartoittamisprosessi myös vastaa paremmin toimitusketjujen dynaamisuuteen, koska automaattinen prosessi on manuaalista prosessia nopeampi viedä läpi. Automatisoimalla kartoittamisprosessi toimitusketjukarttojen reaaliaikaisuus ja päivitettävyyys parantuvat siis merkittävästi.

Nämä kaikki yhdessä mahdollistavat perinteistä kartoittamista paremman skaalautuvuuden. Kielitaito mahdollistaa keskittymisen jollekin tietylle alueelle, josta löytyy tietoa eniten paikallisella kielellä mutta se mahdollistaa myös globaalien toimitusketjujen kartoittamisen, koska parhaimmassa tilanteessa millä tahansa luonnollisella kielellä kirjoitettuja tekstilähteitä voidaan hyödyntää.

Kustannustehokkuus edistää globaalien toimitusketjujen kartoittamista, joiden perinteinen kartoittaminen on hyvin resurssivaativaa mutta alhaisemmat kartoittamisen kustannukset mahdollistavat myös useamman mittakaavaltaan pienempien karttojen luomisen. Automaattisen prosessin nopeus myös tukee useamman kartan luontia, koska

yhden kartan tuottamiseen vaadittu aika vähentyy merkittävästi perinteiseen kartoittamiseen verrattuna.

Luonnollisen kielen käsittelyyn perustuvassa prosessissa aiheutuu kuitenkin myös kustannuksia esimerkiksi esimerkkiaineiston annotoinnissa ja korpuksen laatimisessa, joten prosessin automatisoinnin kustannushyötyä tulee arvioida. Huomautettavaa kuitenkin on, että nämä hyödyt eivät ole vielä nykyteknologialla saavutettavissa ja ne ovat riippuvaisia teknologian kehityksestä tulevaisuudessa.

5 Yhteenveto ja johtopäätökset

Tutkielmassa tarkasteltiin luonnollisen kielen käsittelyyn perustuvan teknologian kykyä luoda automaattisesti toimitusketjukarttoja julkisesti saatavilla olevista tekstilähteistä. Luonnollisen kielen käsittelyn hyödyntämistä toimitusketjujen kartoittamisessa tarkasteltiin tarkemmin luvussa 4.

Luvussa 2 toimitusketjujen kartoittamisen todettiin olevan keino visualisoida yrityksen toimitusketjua sen toiminnan paremmaksi ymmärtämiseksi (Barroso ym. 2011, 169–173; Gardner & Cooper 2003, 37, 39). Toimitusketjun kartoittamisella tunnistettiin olevan selkeitä hyötyjä liiketoiminnassa mutta sen tunnistettiin olevan myös hyvin haastava prosessi. Toimitusketjukartoista tuotiin esille, mitä niiden tulisi sisältää ja miten toimitusketjun eri osia kartoitetaan.

Luvussa 3 tuotiin esille luonnollisen kielen käsittelyn teknologiaa. Luonnollisen kielen käsittelyn todettiin olevan ihmisten väliseen kommunikaation tarkoitetun kielen muuttamista tietokoneen ymmärtämään muotoon ja tietokoneen tekemää päättelyä tekstin sisällöstä (Li ym. 2021, 570; Khurana ym. 2023, 3715–3720). Luonnollisen kielen käsittelystä tuotiin esille sen käytännön prosessi ja se, miten tietokoneen luonnollisen kielen käsittelyssä luoma tieto voidaan ontologian avulla visualisoida toimitusketjukartaksi.

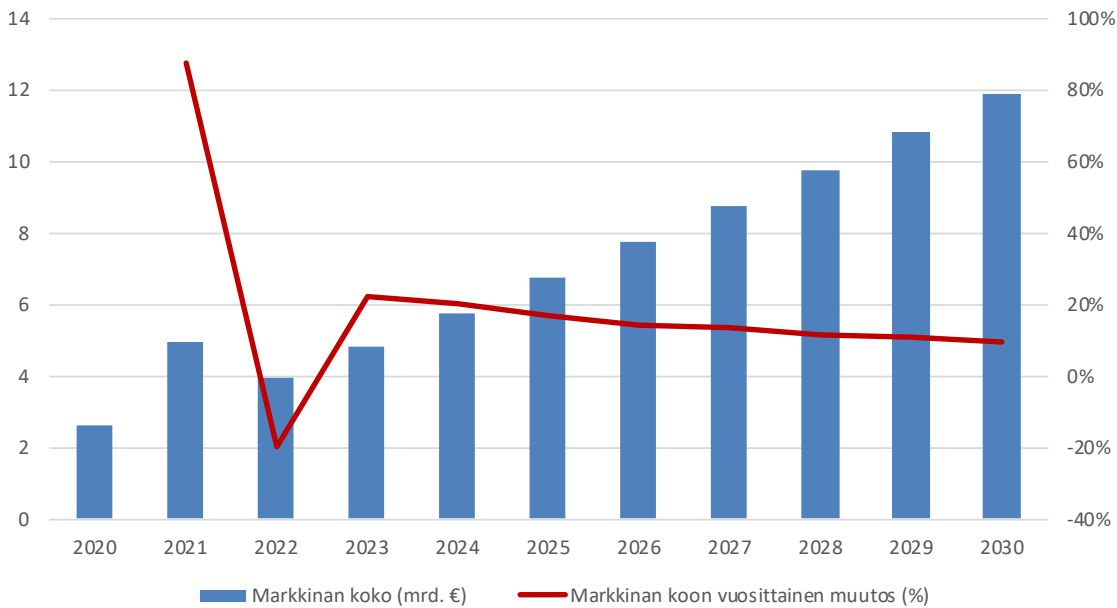
Ensimmäiseen tutkimuskysymykseen ”Mitkä luonnollisen kielen käsittelyn menetelmät soveltuvat hyödynnettäväksi toimitusketjujen kartoittamisessa?” vastattiin luvussa 4.1. Luonnollisen kielen käsittelyn käyttökohteista toimitusketjujen kartoittamiseen tärkein on tiedonpoiminta, jonka tarkoituksena on löytää rakenteettomista tekstilähteistä haluttua tietoa ja muuttaa se rakenteelliseen muotoon (Cowie & Lehnert 1996, 81; Piskorski & Yangarber 2013, 24). Tähän mennessä luodut automaattisen kartoittamisen menetelmät pystyvät hyödyntämään tiedonpoiminnassa nimien tunnistamista ja yhteyksien poimintaa. Tulevaisuudessa menetelmiä on tarpeen kehittää, jotta ohjelma voi hyödyntää myös tapahtumien poimintaa ja samaviitteisyyden tunnistamista. (Wichmann ym. 2020, Wichmann 2022.) Toimitusketjujen kartoittamista varten soveltuvimman luonnollisen kielen käsittelyyn perustuvan ohjelman todettiin olevan koneoppiva muuntajan toimintaan perustuva luokittelija, jonka tehtävänä on tunnistaa, jos raakamuotoisesta tekstilähteestä löytyy toimittajasuhde (Wichmann ym. 2020, 5325;

Wichmann 2022, 100). Wichmann ym. (2020) ja Wichmann (2022) toteuttamat luokittelijat perustuivat vielä muuntajaa yksinkertaisempaan kaksisuuntaiseen takaisinkytkettyyn neuroverkkoon, joten tulevaisuudessa on tarpeellista toteuttaa muuntajan toimintaan perustuva luokittelija. Tulevaisuudessa luokittelijan sisältämien luokkien määrää tulee myös lisätä, jotta sillä voidaan toteuttaa monimutkaisempaa toimitusketjujen kartoittamista ja vastata paremmin luvussa 4.2 esitettyihin haasteisiin (Wichmann ym. 2020, 5332–5333). Luokittelijan monikielisyyttä tulee myös kehittää, jotta luokittelijan avulla hyödynnettävissä olevan datan määrä kasvaa (Wichmann 2022, 56).

Toiseen tutkimuskysymykseen ”Mitä edellytyksiä yrityksiltä vaaditaan toimitusketjujen kartoittamisen automatisoimiseksi luonnollisen kielen käsittelyn avulla?” vastattiin luvussa 4.3. Toimitusketjujen kartoittamisen automatisointi lähtee liikkeelle siitä, että yritys toteuttaa luokittelijana toimivan ohjelman käytännössä. Yleisluontoisissa tekstilähteissä toimijasuhteiden esiintyvyys on melko alhaista (Hiltunen ym. 2017; Wichmann ym. 2020, 5326), minkä takia yrityksen tulee kerätä itselleen tekstilähteitä ohjelmassa käytettäviä korpuksia varten. Tekstilähteitä voidaan kerätä esimerkiksi suodattamalla tunnetun toimitusketjun toimijan mukaan (Wichmann ym. 2022, 165). Luokittelijan kouluttamiseksi osalle kerätyistä tekstilähteistä tulee luoda annotaatioiden avulla esimerkkejä tunnistettavista luokista ja näiden avulla luokittelijaa tunnistamaan luokkien ominaisuuksia tekstilähteistä. Yritys voi ulkoistaa annotaatioprosessin ulkoiselle toimijalle tai toteuttaa sen itse. Jos yritys päättää toteuttaa annotaation itse, yrityksen tulee prosessia varten palkata projektijohtaja, kuraattori, sekä useampia annotaattoreita, sekä hankkia projektin käyttöön annotaatio-ohjelmisto.

Kolmanteen tutkimuskysymykseen ”Mikä on automatisoidun kartoittamisen hyöty verrattuna perinteiseen kartoittamiseen?” vastattiin luvussa 4.4. Automatisoidun prosessin suurin hyöty on sen kustannustehokkuus verrattuna perinteiseen kartoittamiseen. Automaattisesti toteutettu kartoittamisprosessi on perinteistä kartoittamisprosessia merkittävästi nopeampi viedä läpi, mikä parantaa luotujen toimitusketjukarttojen päivitettävyyttä ja reaaliaikaisuutta. Automaattisessa kartoittamisprosessissa on myös merkittävästi enemmän dataa hyödynnettävissä tiedonpoiminnan hyödyntämisen ja kielimallin ihmisiin verrattuna paremman kielitaidon avulla. Nämä kaikki yhdessä mahdollistavat toimitusketjukarttojen paremman skaalautuvuuden, globaalien toimitusketjukarttojen luominen ei ole

perinteiseen kartoittamiseen verrattuna yhtä resurssiavaavaa ja nopeampi kartoittamisprosessi mahdollistaa useamman pienemmän mittakaavan toimitusketjukartan toteuttamisen. Automaattisen kartoittamisen saavuttamia hyötyjä tulee myös puntaroida sen aiheuttamien kustannusten valossa. Huomattavaa kuitenkin on, ettei nykyteknologialla pystytä vielä saavuttamaan mainittuja hyötyjä ja niiden realisoituminen on riippuvaista teknologian kehityksestä tulevaisuudessa.



Kuvio 12: Tekstiperustaisen luonnollisen kielen käsittelyn markkinan koon muutos vuosina 2020–2030 (Statista 2023)

Tekstiin perustuvan luonnollisen kielen käsittelyn markkina tulee kasvamaan tulevaisuudessa tasaisesti (Kuvio 12), joten edellytykset siihen liittyvän teknologian kehittämiseen ovat hyvät myös tulevaisuudessa. Tekstiperustaisen luonnollisen kielen käsittelyn markkinan ja toimitusketjujen läpinäkyvyyden merkityksen kasvu (Kuvio 1; Mubarik ym. 2023, 2653) viittaavat vahvasti siihen, että aihealueeseen liittyvä tutkimus tulee saamaan yhä enemmän huomiota ja rahoitusta tulevaisuudessa. Jotta teknologian laajempi käyttöönotto yrityksissä on mahdollista tulevaisuudessa, tulee tulevaisuuden kehityksen pystyä vastaamaan luvussa 4.2 esiteltyihin haasteisiin. Aihealueen tutkimuksen jatkaminen tulevaisuudessa on siis välttämätöntä, jotta teknologian laajempi käyttöönotto on mahdollista.

Lähteet

- Al-Khatib, Ayman Wael (2023) The impact of industrial Internet of things on sustainable performance: the indirect effect of supply chain visibility. *Business Process Management Journal* 29(5), 1607–1629
- Aronoff, Mark – Rees-Miller, Janie (2017) *The handbook of linguistics*. 2nd edition. Wiley-Blackwell, Chichester, West Sussex, England
- Banko, Michele – Brill, Eric (2001) Scaling to Very Very Large Corpora for Natural Language Disambiguation. *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 26–33
- Barratt, Mark – Oke, Adekoge (2007) Antecedents of supply chain visibility in retail supply chains: A resource-based theory perspective. *Journal of Operations Management* 25(6), 1217–1233
- Barroso, A. P. – Machado, V. H. – Cruz Machado, V. (2011) Supply Chain Resilience Using the Mapping Approach. Teoksessa Li, Pengzhong (toim.), *Supply Chain Management*, 161–184. 1st Edition. IntechOpen, Croatia
- Basile, Alessandro – Crupi, Riccardo – Grasso, Michele – Mercanti, Alessandro – Regoli, Daniele – Scarsi, Simone – Yang, Shuyi – Cosentini, Andrea Claudio (2024) Disambiguation of company names via deep recurrent networks. *Expert Systems with Application* 238, 122035
- Benzi, Francesco – Clava, Riccardo – Bassi, Ezio (2018) Increasing the value stream mapping potential in an industrial process, with a dynamic model, based on data from an industrial ethernet bus. *ECEEE Industrial Summer Study 2018*, 189–195
- Biemann, Chris – Bontcheva, Kalina – Eckart de Castilho, Richard – Gurevych, Iryna – Muhie Yimam, Seid (2017) Collaborative Web-Based Tools for Multi-layer Text Annotation. Teoksessa Ide, Nancy – Pustejovsky, James (toim.), *Handbook of Linguistic Annotation*, 229–256. 1st Edition. Springer Netherlands
- Brintrup, Alexandra (2010) Behaviour adaptation in the multi-agent, multi-objective and multi-role supply chain. *Computers in Industry* 61(7), 636–645
- Bullini Orlandi, Ludovico – Pierce, Paul (2020) Analysis or intuition? Reframing the decision-making styles debate in technological settings. *Management Decision* 58(1), 129–145
- Business Continuity Institute, The (2023) *BCI Supply Chain Resilience Report 2023*. Viitattu 5.2.2024 <https://www.sgs.com/en/>

</media/sgscorp/documents/corporate/brochures/sgs-kn-bci-supply-chain-resilience-report-2023-sponsored-by-sgs-en.cdn.en.pdf>

- Cardoso, Jefferson Rosa – Pereira, Ligia Maxwell – Iversen, Maura Daly – Ramos, Adilson Luiz (2014) What is gold standard and what is ground truth? *Dental Press Journal of Orthodontics* 19(5), 27–30
- Christopher, Martin – Towill, Denis R. (2002) Developing Market Specific Supply Chain Strategies. *The International Journal of Logistics Management* 13(1), 1–14
- Christopher, Martin – Peck, Helen (2004) Building the resilient supply chain. *The International Journal of Logistics Management* 15(2), 1–13
- Chomsky, Noah (1956) Three models for the description of language. *I.R.E. Transactions on Information Theory* 2(3), 113–124
- Cooper, Martha C. – Ellram, Lisa M. – Gardner, John T. – Hanks, Albert M. (1997) Meshing multiple alliances. *Journal of Business Logistics* 18(1), 67–89
- Cowie, Jim – Lehnert, Wendy (1996) Information extraction. *Communications of the ACM* 39(1), 80–91
- Craighead, Christopher W. – Blackhurst, Jennifer – Rungtusanatham, M. Johnny – Handfield, Robert B. (2007) The Severity of Supply Chain Disruptions: Design Characteristics and Mitigation Capabilities. *Decision Sciences* 38(1), 131–156
- Dale, Robert (2010) Classical Approaches to Natural Language Processing. Teoksessa Indurkha, Nitin – Damerau, Frederick Jacob (toim.), *Handbook of natural language processing*, 1–7. 2nd edition. Taylor & Francis, Boca Raton
- Deng, Li – Liu, Yang (2018) A Joint Introduction to Natural Language Processing and to Deep Learning. Teoksessa Deng, Li – Liu, Yang (toim.), *Deep Learning in Natural Language Processing*, 1–22. 1st Edition. Springer Nature Singapore Pte Ltd, Singapore.
- Denis, Christophe – Dion-Blanc, Charlotte – Ella-Mintsa, Eddy – Tran, Viet Chi (2024) Nonparametric plug-in classifier for multiclass classification of S.D.E. paths. *Scandinavian Journal of Statistics* 2024, 1–58
- Dubois, Anna – Gadde, Lars-Erik – Mattson, Lars-Gunnar (2021) Purchasing behaviour and supplier base evolution – a longitudinal case study. *The Journal of Business & Industrial Marketing* 36(5), 689–705
- Eisenstein, Jacob (2019) *Introduction to Natural Language Processing*. 1st Edition. MIT Press, Cambridge, Massachusetts

- Evert, Stefan – Greiner, Paul – Baigger, João Filipe – Lang, Bastian (2016) A distributional approach to open questions in market research. *Computers in Industry* 78, 16–28
- Fabbe-Costes, Nathalie – Lechaptois, Lucie – Spring, Martin (2020) “The map is not the territory”: a boundary objects perspective on supply chain mapping. *International Journal of Operations & Production Management* 40(9), 1475–1497
- Farris, Theodore (2010) Solutions to strategic supply chain mapping issues. *International Journal of Physical Distribution & Logistics Management* 40(3), 164–180
- Farris, Theodore – Ahmad, Fayez (2019) Mapping and managing the financial supply chain. Teoksessa Tate, Wendy – Bals, Lydia – Ellram, Lisa M. (toim.), *Supply Chain Finance: Risk Management, Resilience and Supplier Management*, 139–159. 1st Edition. Kogan Page Limited.
- Finlayson, Mark A. – Erjavec, Tomaž (2017) Overview of Annotation Creation: Processes and Tools. Teoksessa Ide, Nancy – Pustejovsky, James (toim.), *Handbook of Linguistic Annotation*, 167–192. 1st Edition. Springer Netherlands
- Gansler, Jacques S. – Luby, Robert E. Jr. – Kornberg, Bonnie (2004) What is Supply Chain Management? Teoksessa Gansler, Jacques S. – Luby, Robert E. Jr. (toim.), *Transforming Government Supply Chain Management*, 1–18. 1st Edition. Rowman & Littlefield, Lanham, Maryland
- Gardner, John T. – Cooper, Martha C. (2003) Strategic supply chain mapping approaches. *Journal of Business Logistics* 24(2), 37–64
- Gharehchopogh, F. S. – Khalifelu, Z. A. (2011) Analysis and Evaluation of Unstructured Data: Text Mining versus Natural Language Processing. 5th International Conference on Application of Information and Communication Technologies, 1–4
- Giacomello, Giampiero – Preka, Oltion (2023) Sources of strength: mapping the defence sector in Europe. *Defence Studies* 23(4), 531–560
- Goodwin, Geoffrey P. – Johnson-Laird, P.N. (2008) Transitive and pseudo-transitive inferences. *Cognition* 108(2), 320–352
- Green, Forrest B. (2001) Managing the unmanageable: integrating the supply chain with new developments in software. *Supply Chain Management* 6(5), 208–211

- Grefenstette, Gregory – Tapanainen, Pasi (1994) What is a word, What is a sentence? Problems of Tokenization.
- Grosman, Jonatas S. – Furtado, Pedro H.T. – Rodrigues, Ariane M.B. – Schardong, Guilherme G. – Barbosa, Simone D.J. – Lopes, Hélio C.V. (2020) Eras: Improving the quality control in the annotation process for Natural Language Processing tasks. *Information Systems* 93, 101553
- Gruber, Thomas R. (1993) A translation approach to portable ontology specification. *Knowledge Acquisition* 5(2), 199–220
- Halliday, M.A.K. – Yallop, Collin (2007) *Lexicology: A short introduction*. 1st Edition. Continuum, London; New York
- Hiltunen, Turo – McVeigh, Joe – Säily, Tanja (2017) Big and Rich Data in English Corpus Linguistics: Methods and Explorations. *Studies in Variation, Contacts and Change in English* 19
- Hines, Peter – Rich, Nick (1997) The seven value stream mapping tools. *International Journal of Operations and Production Management* 17(1), 46–64
- Hosseinia Shavaki, Fahimeh – Ebrahimi Ghahnavieh, Ali (2022) Applications of deep learning into supply chain management: a systematic literature review and a framework for future research. *Artificial Intelligence Review* 56(5), 4447–4489
- Huang, Guang-Bin – Zhou, Hongming – Ding, Xiaojian – Zhang, Rui (2012) Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics – Part B, Cybernetics* 42(2), 513–529
- Huang, Zhuoyu – Kim, Jiwon – Sadri, Alireza – Dowey, Steve – Dargusch, Matthew S. (2019) Industry 4.0: Development of a multi-agent system for dynamic value stream T mapping in SMEs. *Journal of Manufacturing Systems* 52, 1–12
- Hussain, Rizwana Shaheen – Ruikar, Kirti – Enoch, Marcus P – Brien, Nigel – Gartside, David (2011) Process mapping for road works planning and coordination. *Built environment project and asset management* 7(2), 157–172
- Inomata, Satoshi – Hanaka, Tesshu (2024) Measuring exposure to network concentration risk in global supply chains: Volume versus frequency. *Structural Change and Economic Dynamics* 68, 177–193
- Ittoo, Ashwin – Nguyen, Le Minh – van den Bosch, Antal (2016) Text analytics in industry: Challenges, desiderata and trends. *Computers in Industry* 78, 96–107

- Ivanov, Dmitry (2020) Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transportation Research Part E: Logistics and Transportation Review* 136, 101922
- Ivanov, Dmitry – Dolgui, Alexandre (2020) Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. *International Journal of Production Research* 58(10), 2904–2915
- Janiesch, Christian – Zschech, Patrick – Heinrich, Kai (2021) Machine Learning and Deep Learning. *Electronic Markets* 31(3), 685–695
- Kandasamy, Saravanakumar – Cherukuri, Aswani Kumar (2020) Query expansion using named entity disambiguation for a question-answering system. *Concurrency and Computation* 32(4), 5119
- Katamba, Francis (1989) *An Introduction to Phonology*. 1st Edition. Longman Group UK Limited, London
- Kazemi, Yasaman – Szmerekovsky, Joseph (2015) Modeling downstream petroleum supply chain: The importance of multi-mode transportation to strategic planning. *Transportation Research Part E: Logistics and Transportation Review* 83, 111–125
- Khurana, Diksha – Koli, Aditya – Khatter, Kiran – Singh, Sukhdev (2023) Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications* 82(3), 3713–3744
- Klotz, Leidy – Horman, Michael – Bi, Henry H. – Bechtel, John (2008) The impact of process mapping on transparency. *International Journal of Productivity and Performance Management* 57(8), 623–636
- Kumar, B. Sharma (2016) Value Stream Mapping-A Lean Manufacturing Approach to Reduce the Process Wastages in Clothing Industry. *International Journal of Advances in Management and Economics* 5(5), 23–33
- Lambert, Douglas M. – Cooper, Martha C. – Pagh, Janus D. (1998) Supply Chain Management: Implementation Issues and Research Opportunities. *The International Journal of Logistics Management* 9(2), 1–20
- Ledoux, Kerry – Gordon, Peter C. – Camblin, C. Christine – Swaab, Tamara Y. (2007) Coreference and lexical repetition: Mechanisms of discourse integration. *Memory & Cognition* 35(4), 801–815

- Li, Yan – Thomas, Manoj a – Liu, Dapeng (2021) From semantics to pragmatics: where IS can lead in Natural Language Processing (NLP) research. *European Journal of Information Systems* 30(5), 569–590
- Li, Xun – Wu, Qun – Goldsby, Thomas J. – Holsapple, Clyde W. (2022) Enduring buyer–supplier relationship and buyer performance: the mediating role of buyer–supplier dyadic embeddedness and supplier external embeddedness. *European Journal of Management Studies* 27(3), 291–316
- Litschko, Robert – Vulić, Ivan – Ponzetto, Simone Paolo – Glavaš, Goran (2022) On cross-lingual retrieval with multilingual text encoders. *Information Retrieval Journal* 25(2), 149–183
- Liu, James N. K. – He, Yu-Lin – Lim, Edward H. Y. – Wang, Xi-Zhao (2013) A New Method for Knowledge and Information Management Domain Ontology Graph Model. *IEEE Transaction on Systems, Man and Cybernetics: Systems* 43(1), 115–127
- Lyon, Caroline – Ray, Frank (1998) Using Single Layer Networks for Discrete, Sequential Data: An Example from Natural Language Processing. *Neural Computing & Applications* 5, 196–214
- MacCarthy, Bart L. – Ahmed, Wafaa A. H. – Demirel, Guven (2022) Mapping the supply chain: Why, what and how? *International Journal of Production Economics* 270, 108688
- Maedche, Alexander – Staab, Steffen (2001) Ontology learning for the Semantic Web. *IEEE Intelligent Systems* 16(2), 72–79
- Matthews, Peter Hugoe (1991) *Morphology*. 2nd Edition. Cambridge University Press, Cambridge
- Mefford, Robert N. (2011) The Economic Value of a Sustainable Supply Chain. *Business and Society Review* 116(1), 109–143
- Mehrabi, Ninareh – Morstatter, Fred – Saxena, Nripsuta – Lerman, Kristina – Galstyan, Aram (2021) A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys* 54(6), Article 115, 1–35
- Mubarik, Muhammad Shujaat – Naghavi, Navaz – Mubarik, Movashar – Kusi-Sarpong, Simonov – Khan, Sharfuddin Ahmed – Zaman, Syed Imran – Kazmi, Syed Hasnain Alam (2021) Resilience and cleaner production in industry 4.0: Role of supply chain mapping and visibility. *Journal of Cleaner Production* 292, 1–12

- Mubarik, Muhammad Shujaat – Simonov, Kusi-Sarpong – Govindan, Kannan – Khan, Sharfuddin Ahmed – Oyedijo, Adegboyega (2023) Supply chain mapping: a proposed construct. *International Journal of Production Research* 61(8), 2653–2669
- Müller, Christopher – Strube, Michael (2006) Multi-level annotation of linguistic data with MMAX2. Teoksessa Braun, Sabine – Kohn, Kurt – Mukherjee, Jobrato (toim.) *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, 197–214. 1st Edition. Lang, Germany
- Nadeau, David – Sekine, Satoshi (2007) A survey of named entity recognition and classification. *Linguisticae Investigationes* 30(1), 3–26
- Nadkarni, Prakash M. – Ohno-Machado, Lucila – Chapman, Wendy W. (2011) Natural language processing: an introduction. *Journal of the American Medical Informatics Association* 18(5), 544–551
- Newman, John – Cox, Christopher (2020) *Corpus Annotation*. Teoksessa Paquot, Magali – Gries, Stefan Thomas (toim.), *A practical handbook of corpus linguistics*, 25–48. 1st Edition. Springer
- Norwood, F. Bailey – Peel, Derrell (2021) Supply Chain Mapping to Prepare for Future Pandemics. *Applied Economic Perspectives and Policy* 43(1), 412–429
- Orasan, Constantin – Ha, Le An – Evans, Richard – Hasler, Laura – Mitkov, Ruslan (2007) Corpora for computational linguistics. *Ilha Do Desterro* (52), 65–101
- Palmer, David D. (2010) Text Preprocessing. Teoksessa Indurkha, Nitin – Damerau, Frederick Jacob (toim.), *Handbook of natural language processing*, 9–30. 2nd edition. Boca Raton, Taylor & Francis
- Peng, Yaohao – Nagata, Mateus Hiro (2020) An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data. *Chaos, Solitons and Fractals* 139, 110055
- Piskorski, Jakub – Yangarber, Roman (2013) *Information Extraction: Past, Present and Future*. Teoksessa Poibeau, Terry (toim.), *Multi-source, multilingual information extraction and summarization*, 23–49. 1st Edition. Berlin, Springer
- Ponomarov, Serhiy Y. – Holcomb, Mary C. (2009) Understanding the concept of supply chain resilience. *The International Journal of Logistics Management* 20(1), 124–143
- Popat, Kashyap – Mukherjee, Subhabrata – Strötgen, Jannik – Weikum, Gerhard (2016) Credibility assessment of textual claims on the web. *Proceedings of the 25th*

ACM International on Conference on Information and Knowledge Management, 2173–2178

Pustejovsky, James – Stubbs, Amber (2012) *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. 1st Edition. O'Reilly Media Inc, Sebastopol, California

Pustejovsky, James – Bunt, Harry – Zaenen, Annie (2017) *Designing Annotation Schemes: From Theory To Model*. Teoksessa Ide, Nancy – Pustejovsky, James (toim.), *Handbook of Linguistic Annotation*, 21–72. 1st Edition. Springer Netherlands

Roodschild, Matías – Gotay Sardiñas, Jorge – Will, Adrián (2020) A new approach for the vanishing gradient problem on sigmoid activation. *Progress in Artificial Intelligence* 9(4), 351–360

Rutowski, Tomek – Harati, Amir – Shriberg, Elizabeth – Lu, Yang – Piotr, Chlebek – Oliveira, Ricardo (2022) Toward Corpus Size Requirements for Training and Evaluating Depression Risk Models Using Spoken Language. 23rd INTERSPEECH Conference, 3343–3347

Saeed, John I. (2016) *Semantics*. 4th Edition. Wiley-Blackwell, Chichester, West Sussex, England

Schuster, Mike – Paliwal, Kuldip K. (1998) Bidirectional Recurrent Neural Networks. *IEEE Transactions on Signal Processing* 45(11), 2673–2681

Schöpfer, Henning – Kersten, Wolfgang (2021) Using Natural Language Processing for Supply Chain Mapping: A Systematic Review of Current Approaches. 5th International Conference on Computational Linguistics and Intelligent Systems, 71–86

Sharma, Richa – Agarwal, Pooja – Arya, Arti (2022) *Natural Language Processing and Big Data: A Strapping Combination*. Teoksessa Sharma, Rohit – Sharma, Dilip (toim.), *New Trends and Applications in Internet of Things (IoT) and Big Data Analytics*, 255–271. Intelligent Systems Reference Library 221

Simchi-Levi, David – Schmidt, William – Wei, Yehua (2014) From Superstorms to Factory Fires: Managing Unpredictable Supply-Chain Disruptions. *Harvard Business Review* 92(1,2), 96–101

Singh, Jasmeet – Gupta, Vishal (2017) Text Stemming: Approaches, Applications, and Challenges. *ACM Computing Surveys* 49(3), Article 45, 1–46

- Smirnova, Alisa – Cudré-Mauroux, Philippe (2019) Relation Extraction Using Distant Supervision: A Survey. *ACM Computing Surveys* 51(5), Article 106, 1–35
- Snow, Rion – O’Connor, Brendan – Jurafsky, Daniel – Ng, Andrew Y. (2008) Cheap and Fast — But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 254–263
- Songkajorn, Yaninee – Aujiropongpan, Somnuk – Jiraphanumes, Kritsakorn – Pattanasing, Kanittha (2022) Organizational strategic intuition for high performance: The role of knowledge-based dynamic capabilities and digital transformation. *Journal of Open Innovation* 8(3), 1–19
- Sourcemap (2020) Green Wolf Distribution. Viitattu 21.3.2024
<https://open.sourcemap.com/maps/58dc588d31373626640e347f>
- Statista (2018) Biggest supply chain challenges worldwide 2017-2018. Viitattu 21.1.2024 <https://www-statista-com.ezproxy.utu.fi/statistics/829634/biggest-challenges-supply-chain/>
- Statista (2023) Text-based NLP. Viitattu 6.4.2024 <https://www-statista-com.ezproxy.utu.fi/outlook/tmo/artificial-intelligence/natural-language-processing/text-based-nlp/worldwide>
- Säily, Tanja – Tyrkkö, Jukka (2021) Challenges of combining structured and unstructured data in corpus development. *Research in Corpus Linguistics* 9(1), 1–8
- Tabassum, Ayisha – Patil, Rajendra R. (2020) A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing. *International Research Journal of Engineering and Technology* 7(6), 4864–4867
- Thieberger, Nick – Margetts, Anna – Morey, Stephen – Musgrave, Simon (2016) Assessing Annotated Corpora as Research Output. *Australian Journal of Linguistics* 36(1), 1–21
- Traxler, Matthew J. (2011) Parsing. *Wiley Interdisciplinary Reviews* 2(4), 353–364
- Tuo, Meimei – Yang, Wenzhong (2023) Review of entity relation extraction. *Journal of Intelligent & Fuzzy Systems* 44(5), 7391–7405
- Vaswani, Ashish – Shazeer, Noam – Parmar, Niki – Uszkoreit, Jakob – Jones, Llion – Gomez, Aidan N. – Łukasz, Kaiser – Polosukhin, Illia (2017) Attention Is All You Need. *31st Conference on Neural Information Processing Systems*, 6000–6010

- Weerakkody, Vishanth – Janssen, Marjin – Dwivedi, Yogesh K. (2011)
Transformational change and business process reengineering (BPR): Lessons from the British and Dutch public sector. *Government Information Quarterly* 28(3), 320–328
- Wichmann, Pascal – Brintrup, Alexandra – Baker, Simon – Woodall, Philip – McFarlane, Duncan (2018) Towards automatically generating supply chain maps from natural language text. *IFAC PapersOnLine* 51(11), 1726–1731
- Wichmann, Pascal – Brintrup, Alexandra – Baker, Simon – Woodall, Philip – McFarlane, Duncan (2020) Extracting supply chain maps from news articles using deep neural networks. *International Journal of Production Research* 58(17), 5320–5336
- Wichmann, Pascal (2022) Towards automatically generating supply chain maps from natural language text. University of Cambridge, Department of Engineering
- Xiang, Wei – Wang, Bang (2019) A Survey of Event Extraction From Text. *IEEE Access* 7, 173111–173137
- Zaheer, Manzil – Guruganesh, Guru – Dubey, Avinava – Ainslie, Joshua – Alberti, Chris – Ontanon, Santiago – Pham, Philip – Ravula, Anirudh – Wang, Qifan – Yang, Li – Ahmed, Amr (2020) Big Bird: Transformers for Longer Sequences. 34th Conference on Neural Information Processing Systems, 17283–17298
- Zhou, Deyu – He, Yulan (2008) A hybrid generative/discriminative framework to train a semantic parser from an un-annotated corpus. The 22nd International Conference on Computational Linguistics, 1113-1120